

Politechnika Śląska w Gliwicach
Wydział Automatyki, Elektroniki i Informatyki

Rozprawa doktorska

mgr inż. Marek Drewniak



**Politechnika
Śląska**

**METODY ANALIZY EFEKTYWNOŚCI I JAKOŚCI PROCESÓW
PRODUKCJI DYSKRETNEJ
WYKORZYSTUJĄCE WZORCE TECHNOLOGICZNE
I WYBRANE MECHANIZMY EKSPLOKACJI DANYCH**

Promotor:

dr hab. inż. Rafał Cupek

*Ze szczególnymi podziękowaniami
dla mojej Żony
za cierpliwość i motywowanie podczas pisania niniejszej pracy*

oraz

*dla Pana dr hab. inż. Rafała Cupka
za inspirację do prowadzenia badań i wszelką pomoc*

SPIS TREŚCI

1. WSTĘP.....	7
1.1. Wprowadzenie	7
1.2. Motywacja do podejmowanych działań.....	10
1.3. Cel i zakres pracy	13
1.4. Opis ogólny prowadzonych badań.....	15
1.5. Zawartość pracy	19
2. ANALIZA DZIEDZINY I PRZEGLĄD ISTNIEJĄCYCH ROZWIĄZAŃ.....	20
2.1. Systemy realizacji produkcji.....	20
2.2. Modelowanie usług i danych produkcyjnych z wykorzystaniem podejścia obiektowego	29
2.3. Komunikacja w przemysłowych systemach informatycznych wykorzystująca modele obiektowe	40
2.4. Analiza skupień w eksploracji danych.....	46
2.4.1. Klasteryzacja metodą <i>k</i> -Means	53
2.5. Detekcja wariantów i profili produkcyjnych	58
2.6. Klasteryzacja danych zbieranych w przemysłowych systemach informatyki	64
2.7. Wykrywanie anomalii produkcyjnych z wykorzystaniem technik eksploracji danych.....	68
3. BADANIA NAD ZASTOSOWANIEM MODELI INFORMACYJNYCH DO WSPARCIA UTRZYMANIA INSTALACJI PRZEMYSŁOWYCH	77
3.1. Model informacyjny technologii instalacji produkcyjnej	77
3.2. Model informacyjny wzorców technologicznych.....	81
3.3. Integracja danych procesowych z wiedzą technologiczną.....	84
3.4. Podsumowanie	90

4. METODA AUTOMATYCZNEGO WYKRYWANIA WZORCÓW TECHNOLOGICZNYCH Z WYKORZYSTANIEM KLASTERYZACJI METODĄ <i>k</i> -MEANS	92
4.1. Detekcja cykli produkcyjnych i stanów znaczących wykorzystywanych w wykrywaniu wzorców technologicznych.....	98
4.2. Detekcja wzorców technologicznych z wykorzystaniem algorytmu klasteryzacji <i>k</i> -Means.....	107
4.3. Automatyzacja wyznaczania liczby wzorców technologicznych z użyciem zmodyfikowanego algorytmu klasteryzacji <i>k</i> -Means	112
4.4. Podsumowanie	122
5. METODA AUTOMATYCZNEGO WYKRYWANIA ANOMALII W PRACY URZĄDZEŃ.....	125
5.1. Dopasowanie bieżących obserwacji do wyznaczonych wzorców technologicznych	125
5.2. Detekcja i klasyfikacja anomalii w pracy urządzeń.....	129
5.3. Podsumowanie	135
6. ZAKOŃCZENIE	138
7. BIBLIOGRAFIA	143
ZAŁĄCZNIK 1 - OPIS STANOWISKA PRODUKCYJNEGO WYKORZYSTYWANEGO W TRAKCIE BADAŃ	156
Stanowisko DC40	156

1. WSTĘP

1.1. Wprowadzenie

Podejście do produkcji w branżach przemysłowych przechodzi w ostatnich latach gruntowną transformację. Trendy obserwowane w przemyśle wskazują na konieczność wprowadzania zmian nie tylko ze względu na pojawianie się nowych technologii, ale przede wszystkim z powodu zmieniającego się podejścia przedsiębiorstw do wytwarzania. Rynki produkcyjne coraz częściej muszą dostosowywać się do szybkich zmian w wymaganiach produktowych, uwzględniać dużą wariantowość, reagować na nieciągłość łańcuchów dostaw lub na braki dostępności części zamiennych do maszyn. Jednocześnie przemysł dąży do ciągłego poprawiania jakości wytwarzanych wyrobów, zwiększania efektywności procesów technologicznych, minimalizacji przestoju wynikających ze zmian technologicznych i awaryjności urządzeń produkcyjnych oraz zwiększania bezpieczeństwa pracy na terenie zakładów.

Jednym z trendów przemysłowych spopularyzowanych w ostatnich latach jest koncepcja Przemysłu 4.0 (niem. Industrie 4.0). Postępujące zmiany w technikach wytwarzania i problemy, które są z nimi związane, zostały dostrzeżone na początku drugiej dekady XXI wieku przez kilka niezależnych instytucji, między innymi German Platform Industrie-4.0, US National Institute of Standards and Technology czy Smart Manufacturing Leadership Coalition South Korea. Instytucje te doprowadziły do systematyzacji pojęć i pojęć w nowoczesnych technikach wytwarzania, a najbardziej popularnym ich zbiorem stał się właśnie Przemysł 4.0. Oznacza on ogólnie pojętą cyfryzację przemysłu jako następującą, czwartą, po upowszechnieniu maszyny parowej, elektryfikacji i automatyzacji, rewolucję przemysłową. Interpretacji filozofii Przemysłu 4.0 jest tak wiele, jak wielu jest dostawców i producentów różnych technologii stosowanych w przemyśle, jednak ich najważniejszą cechą wspólną jest integracja różnych warstw wykonawczych (urządzeń, systemów, technologii, technik komunikacyjnych) i przenoszenie wiedzy o stosowanych w nich technologiach do postaci cyfrowej tak, aby poprawiać zarządzanie, efektywność, jakość i operacyjność instalacji produkcyjnych. Dzięki integracji systemów kontrolnych automatyki sterującej procesami z systemami informatycznymi, między innymi: do zarządzania i realizacji produkcji (ang. Manufacturing Execution Systems, MES); zarządzania transportem (ang. Transport Management Systems, TMS) czy pracą służb utrzymania ruchu (ang. Computerised

Maintenance Management Systems, CMMS), możliwe jest uzyskanie unikatowego połączenia różnego rodzaju danych. Na danych tych można następnie zastosować techniki eksploracji i modelowania tak, aby odkryć zależności i wzorce zachowań układów wykonawczych niewidoczne bez jednoczesnej analizy danych z wielu źródeł. Efektem tych operacji jest wytworzenie całkowicie nowej wiedzy, umożliwiającej poprawę efektywności, jakości i skuteczności wspomnianych procesów technologicznych. Dodatkowo, dzięki odkrywaniu ukrytych zależności związanych z realizacją procesu, naturalnym staje się wspomaganie pracy personelu odpowiedzialnego za utrzymanie instalacji produkcyjnych, optymalizowanie pracy urządzeń, minimalizacja ryzyka przestoju, zapobieganie występowaniu poważnych awarii, wspomaganie pracy planistów produkcyjnych, działów zaopatrzenia i logistyki oraz redukcja kosztów pracy maszyn i personelu.

Branża produkcyjna charakteryzuje się dość powolnym wdrażaniem innowacji, głównie ze względu na hermetyczność implementowanych rozwiązań technologicznych: raz zaprojektowana i uruchomiona linia technologiczna pracuje zwykle kilka lub kilkanaście lat, podczas których jakiegokolwiek zmiany wdrażane są niechętnie, głównie ze względu na ryzyka związane z ingerencją w pracujące i ustabilizowane układy. Z tego powodu jednostki kontrolne wykorzystywane w procesach technologicznych, m.in. sterowniki swobodnie programowalne PLC (ang. Programmable Logic Controller) czy komputery klasy PC, niejednokrotnie są przestarzałe lub po prostu niedostosowane do możliwości informatycznych, jakie pojawiają się wraz z rozwojem Przemysłu 4.0, tak pod kątem dostępnej mocy obliczeniowej jak dostępnych interfejsów komunikacyjnych zapewniających wystawienie danych procesowych do analizy. Dopiero od kilku lat widoczny jest gwałtowny rozwój sprzętu i mocy obliczeniowych układów kontrolnych automatyki, a dodatkowo obserwowany jest rozwój otwartych standardów komunikacyjnych umożliwiających podłączenie do różnych typów urządzeń bez konieczności stosowania się do specyfiki indywidualnych protokołów i standardów komunikacyjnych. Ponieważ jednak zmiana warstwy sprzętowej następuje powoli, istnieje zaobserwowane przez autora pole do badań, w którym możliwe jest utworzenie zestawu metod analitycznych działających na ograniczonych zasobach i technikach komunikacyjnych, a które przez swoje działanie nie tylko poprawiałyby pracę instalacji, ale przede wszystkim nie ingerowałyby w systemy sterowania w sposób wpływający na bezpieczeństwo i stabilność procesów produkcyjnych.

W transformacji przemysłu nie bez znaczenia pozostaje fakt upowszechnienia i rozwoju technik informatycznych, które przez wiele lat pozostawały w sferze badawczej lub, dla wdrożeń, były nieopłacalne ekonomicznie, głównie ze względu na ograniczone moce obliczeniowe i pamięci kontrolerów procesów, braki w możliwościach komunikacyjnych pomiędzy składowymi instalacji przemysłowych lub ograniczenia występowania systemów informatycznych w wybranych warstwach produkcyjnych. Do technik tych należą metody przetwarzania i przechowywania dużych ilości danych, rozwiązania analityczne wykorzystujące eksplorację danych i sztuczną inteligencję czy wspomaganie zarządzania wiedzą w procesach sterowania. Obecnie, dzięki coraz częstszemu stosowaniu systemów wspomagających produkcję, zastosowaniu komputerów przemysłowych o odpowiedniej mocy obliczeniowej i wyposażeniu urządzeń wykonawczych i pomiarowych automatyki w cyfrowe interfejsy komunikacyjne, możliwe jest skuteczne zintegrowanie danych z maszyn, przetworzenie ich i przekształcenie w wiedzę z wykorzystaniem technik analitycznych analogicznych do tych, które stosowane są w innych dziedzinach, np. finansowej, handlowej lub meteorologicznej. Do najczęściej wykorzystywanych należą metody statystyczne, techniki eksploracji danych, w tym metody uczenia maszynowego, klasteryzacja, wykrywanie asocjacji, wzorców i klasyfikacja. Ich aplikowalność uzależniona jest przede wszystkim od problemu analitycznego do rozwiązania, ale również od specyfiki danych wejściowych oraz złożoności obliczeniowej i dostępnej mocy jednostek przetwarzających. Specyfika produkcji wielowariantowej w naturalny sposób wskazuje na wykorzystanie technik do grupowania danych w celu wykrywania zależności występujących pomiędzy analizowanymi cechami. Z tego powodu techniki analityczne wykorzystane w metodach zaproponowanych przez autora koncentrują się głównie na klasteryzacji. Z kolei powiązanie danych procesowych z fragmentami instalacji i dalsze budowanie wiedzy o procesach realizowane może być w oparciu o modele danych, które w dobie produkcji wielowariantowej wymagają jednak odpowiedniego dostosowania. Aspekt ten stanowił więc dodatkowy, istotny punkt w badaniach i pracy autora.

Dzięki upowszechnianiu się technik komunikacyjnych i możliwości połączenia systemów produkcyjnych o różnej specyfice oraz dostępowi do stosunkowo dużych mocy obliczeniowych i przestrzeni do przechowywania danych do analizy, rynek produkcyjny stopniowo przeobraża się w rynek indywidualizowany dla klienta, w którym wytwarzanie realizowane jest w sposób elastyczny i adaptacyjny, z jednoczesną możliwością

wydobywania informacji i budowania wiedzy w sposób dotychczas mocno ograniczony lub, w wybranych przypadkach, nieosiągalny. Ponieważ jednak zmiany wdrażane są stopniowo, brak jest uniwersalnych narzędzi do spełniania indywidualnych wymagań poszczególnych obiektów produkcyjnych. Dodatkowo, każda tego typu aplikacja cechuje się indywidualnymi wymaganiami i ograniczeniami i istnieje szerokie pole do badań i wdrożenia potencjalnych rezultatów prac badawczych prowadzonych i skupionych wokół gromadzenia i analizowania danych oraz budowania z nich wiedzy. Celem uogólnienia podejścia do analiz, możliwe jest zawężenie obszaru badań do instalacji przemysłowych o wspólnych, uniwersalnych cechach, m.in. rejestrowanemu bieżącemu zużyciu energii zasilających na wejściu, generowaniu danych o aktywności urządzeń wykonawczych przez kontroler procesowy czy generowaniu znaczników czasowych. Dane mogą następnie zostać przekształcone w wartości zagregowane, wykorzystywane w dalszych analizach. Co jednak jest najważniejsze, to fakt, że analizom poddana może zostać dowolna instalacja, bez konieczności posiadania wiedzy na temat jej budowy, sposobu pracy czy wariantów produkcyjnych, jakie realizuje.

Z tego powodu autor rozprawy, mając poprzez pracę zawodową bezpośredni kontakt z rynkiem automatyki przemysłowej, zdecydował o przeprowadzeniu badań nad utworzeniem zestawu metod do analizy danych procesowych z wykorzystaniem wybranych mechanizmów eksploracji danych i powiązania tych mechanizmów z technologią wykonania instalacji. Metody te poprawiać mogą jakość i efektywność pracy instalacji przemysłowych pracujących w produkcji dyskretniej.

1.2. Motywacja do podejmowanych działań

Najważniejszym powodem, dla którego autor niniejszej pracy zdecydował o obranej ścieżce badań, jest słabo wykorzystany potencjał technik analitycznych opartych o metody eksploracji danych, stosowanych dla potrzeb analizy danych produkcyjnych i łączących te analizy z wiedzą technologiczną. Istnieją liczne publikacje i opracowania naukowe, stanowiące o rozwoju tej dziedziny, poruszające wiele problemów, jednak niewyczerpujące zagadnień lub ograniczające stosowanie do wybranych problemów lub specyfik aplikacyjnych. Obejmują one: badania przebiegów wartości zmiennych i szeregów czasowych [1]; wykrywanie wzorców danych dla potrzeb m.in. wykrywania anomalii w pracy urządzeń [2][3][4] ; wykrywanie powiązań pomiędzy wytwarzanymi produktami a problemami jakościowymi [5] lub dla potrzeb analiz wydajnościowych

procesów [6]. Pole do badań pozostaje jednak bardzo szerokie, m.in. ze względu na zmienne specyfiki danych produkcyjnych, trudności w aplikacji rozwiązań, niejednokrotny brak możliwości stosowania opracowanych mechanizmów analitycznych dla wielu różnych przypadków produkcyjnych panujących w jednym środowisku lub brak powiązania analiz z wiedzą o technologii wykonania instalacji. Autor w prowadzonych badaniach położył nacisk na dwa aspekty, które w dotychczasowych opracowaniach pozostawały nierozwiązane lub jedynie wspomniane: na analizie danych z produkcji krótkoseryjnej i wielowariantowej oraz na umożliwieniu efektywnego połączenia analiz z informacjami technologicznymi. Pierwszy przypadek jest nietrywialny ze względu na specyfikę gromadzonych danych, która uniemożliwia naukę na odpowiednio dużych zbiorach wejściowych, a która wynika z braku możliwości lub czasu na odpowiednie ich zgromadzenie. Drugi zakłada opracowanie podejścia, które nie tylko wskaże na korelację pomiędzy wynikami analiz, a składowymi technologicznymi, ale jednocześnie uprości te analizy poprzez wyznaczenie i zastosowanie odpowiednio opracowanych wartości zagregowanych.

Drugim powodem jest dostrzeżenie braku wdrażanych rozwiązań produkcyjnych, które zdolne byłyby do obliczania ogólnych wartości zagregowanych i uniwersalnej analizy danych procesowych i biznesowych, a następnie umożliwiałyby szczegółową analizę tych danych dla wybranych potrzeb, m.in. prewencyjnego i predykcyjnego utrzymania instalacji, wnioskowania na temat jakości prowadzonych procesów lub wyliczania kluczowych wskaźników produkcji. Istnieją oczywiście rozwiązania, m.in. systemy klasy Business Intelligence [7], wspomagające analizy wydajnościowe i podejmowania decyzji, jednak typowo stosowane są one do wsparcia zarządzania całymi obiektami, a nie pojedynczymi instalacjami produkcyjnymi. Autor, pracując zawodowo w firmie technologicznej zajmującej się projektowaniem, uruchamianiem i utrzymaniem instalacji produkcyjnych, posiada kontakt z różnymi typami instalacji i technologiami stosowanymi w przemyśle, a przez to dostrzega zapotrzebowanie na suplementację wiedzy technologicznej i możliwe wpływanie na procesy w celach ich poprawy. Opracowany w ramach pracy zestaw metod analitycznych wpisuje się w zakres badań przemysłowych i może zostać praktycznie wykorzystany, między innymi do wsparcia pracy instalacji.

Trzecim powodem jest zapotrzebowanie na systemy analityczne pracujące w warunkach przemysłowych. Zapotrzebowanie szczególnie widoczne jest w zakładach,

w których oferowane są systemy wspomaganie służb utrzymania ruchu, systemy zarządzania majątkiem przedsiębiorstwa, systemy zarządzania magazynami, transportem i produkcją oraz instalacje automatyki przemysłowej. Bezpośrednio zgłaszanymi potrzebami są m.in. redukcja kosztów przestojów wynikających z awarii, optymalizacja zużycia energii zasilających instalacje i mediów wykorzystywanych podczas produkcji, chęć budowania świadomości o stosowanych technologiach, dążenie do lepszego zrozumienia procesów i wykorzystania takiej wiedzy do lepszego zarządzania pracą zakładu. Dodatkowo, widocznym czynnikiem pozostaje chęć inwestorów do wdrażania rozwiązań dostosowanych do produkcji szybkozmiennnej i krótkoseryjnej, konieczne jest więc, aby systemy wspomagające, takie jak analityczne, dostosowywały się do tej zmienności i adaptowały do nowych warunków pracy, wariantów produkcji i wszelkich cech wynikających ze zmienności procesu. Obecnie brak jest tego typu rozwiązań, co stanowi o dodatkowej motywacji do podejmowanych badań.

Wreszcie, czwartym powodem dla podjęcia prac badawczych przez autora jest jego zainteresowanie dziedziną eksploracji danych, w szczególności technik klasteryzacji, oraz wykorzystanie ich szczególnie w mechanizmach poprawy efektywności energetycznej stanowisk produkcyjnych, m.in. dla potrzeb wykrywania anomalii w pracy, wykrywania postępującej degradacji urządzeń wykonawczych oraz budowania świadomości o prowadzonych procesach produkcyjnych. Zagadnienie cechuje się trudnością na trzech poziomach: poziomie akwizycji, przechowywania i przetwarzania dużej ilości danych o różnym charakterze, napływających z różnych źródeł i z różnymi częstotliwościami; poziomie automatycznej detekcji wybranych parametrów pracy instalacji przy założeniu braku znajomości a priori technologii; poziomie przetwarzania danych o różnym charakterze, z wykorzystaniem technik eksploracji danych i statystycznych.

Zaprezentowane w rozprawie wyniki, w tym wyniki badań realizowanych w warunkach rzeczywistych lub zbliżonych do panujących w rzeczywistych systemach produkcyjnych, wskazują na możliwość poprawy efektywności, jakości i skuteczności pracy instalacji przemysłowych. Skuteczne opracowanie i wdrożenie systemów analitycznych stanowić będzie w niedalekiej przyszłości szeroko stosowany nurt w wielu gałęziach przemysłu, szczególnie wytwórczego. Korzyści z przetwarzania danych procesowych, przy relatywnie niewielkim stopniu ingerencji w istniejącą infrastrukturę produkcyjną oraz niskie koszty wdrożenia, umożliwią lepsze zrozumienie procesów technologicznych, ich ciągłą poprawę oraz być może odsłonią nieznane obecnie problemy

i zagadnienia możliwe do dalszej eksploracji, zarówno w podejściach analizy danych jak i procesach ich gromadzenia, przechowywania lub przekształcania w wiedzę.

1.3. Cel i zakres pracy

Ze względu na objęcie kierunku prac badawczych na produkcję krótkoseryjną i wielowariantową oraz dążąc do powiązania prowadzonych analiz z wiedzą technologiczną, sformułowano następującą tezę główną rozprawy:

Zastosowanie modeli informacyjnych do powiązania danych procesowych z wiedzą technologiczną i wykorzystanie mechanizmów eksploracji danych pozwala na istotną poprawę efektywności pracy i utrzymania dostępności instalacji przemysłowych.

Ponieważ dane procesowe wykorzystywane w standardowych systemach sterowania i wizualizacji uniemożliwiają obserwację zmian i trendów w długich horyzontach czasowych, dla potrzeb analizy krótkich ich serii niemożliwe jest wykorzystanie technik wymagających czasu na gromadzenie. Dodatkowo, duża ilość danych przesyłanych z systemów kontrolnych do systemów analitycznych może stanowić problem przy zyskujących popularność rozwiązaniach bezprzewodowej i mobilnej automatyki, w których układy wykonawcze nie posiadają stałego i stacjonarnego dostępu do sieci komunikacyjnej, np. w przypadku pojazdów AGV (ang. Autonomous Guided Vehicles). Z tego powodu konieczne jest opracowanie metod redukujących złożoność obliczeniową stosowanych algorytmów analitycznych. Podejściem umożliwiającym tą redukcję jest przekształcenie surowych danych wejściowych w wartości zagregowane opisujące wybrane, powtarzalne serie danych produkcyjnych, zdefiniowane jako cykle pracy. Określenie to stosowane jest w przemyśle typowo do oznaczenia czasu i zestawu czynności wymaganego do wyprodukowania pojedynczego produktu w obrębie instalacji przemysłowej. Momenty początku i końca cyklu zależą mogą od stosowanych technologii (np. jako początek uznawać można moment wpuszczenia produktu w obszar stacji produkcyjnej za pomocą siłownika dozującego na przenośniku transportowym lub moment rozpoczęcia pracy specyficznego urządzenia wykonawczego), jednak dla przypadku analitycznego, w którym zakładana jest powtarzalność ze względu na aktywacje, sekwencje pracy, czasy wysterowań urządzeń czy pochłaniane energie, nie posiadają one istotnego znaczenia. Dla rozpatrywanego przypadku analitycznego cykle pracy stanowią wektory cech, zawierające zagregowane wartości opisujące: całkowity czas trwania cyklu,

całkowitą energię zużytą w trakcie cyklu, całkowite czasy wysterowań (aktywacji) poszczególnych urządzeń. Pierwsza z tych dodatkowych sformułowana więc została w postaci:

Agregacja danych wyliczanych w obrębie cykli produkcyjnych pozwala na zredukowanie złożoności obliczeniowej i skrócenie czasu realizacji algorytmów wyszukiwania wzorców w produkcji dyskretnej wielowariantowej.

Produkcja krótkoseryjna i wielowariantowa cechuje się częstym wprowadzaniem nowych typów produktów. Z tego powodu, ze względów utrzymaniowych systemu, utrudnionym jest zasilanie algorytmów analitycznych informacjami o nowowprowadzanych wariantach. Konieczne jest więc automatyczne wykrywanie i oznaczanie nowych typów produktów, tylko na podstawie stałego zestawu danych wejściowych i wyliczanych na ich podstawie wartości zagregowanych. Techniki analityczne muszą więc cechować się adaptacyjnością, tj. wykrywać pojawiające się warianty w trakcie działania, bez konieczności każdorazowego przetrenowywania zbiorów wejściowych. Dodatkowo, założeniem dla projektowanego zestawu technik analitycznych jest ich uniwersalność, niezależnie od stosowanego w analizowanym przypadku medium zasilającego instalację (np. sprężonego powietrza, energii elektrycznej) czy zestawu urządzeń wykonawczych realizujących proces technologiczny. Z tego powodu drugą z tych dodatkowych sformułować można w postaci:

Możliwe jest automatyczne wykrywanie i klasyfikacja profili produkcyjnych instalacji bez znajomości technologii a priori (tj. bez wiedzy na temat stosowanych urządzeń i specyfiki ich pracy, bez znajomości procesu produkcyjnego oraz bez wiedzy na temat aktualnie produkowanego typu produktu), jedynie z wykorzystaniem informacji o zużywanych energiach, czasach trwania i sygnałach binarnych definiujących aktywność urządzeń wykonawczych..

Oprócz wykorzystania rezultatów z analiz do połączenia ich z informacjami technologicznymi i wyprodukowania wiedzy, wyniki te wesprzeć mogą bezpośrednio utrzymanie instalacji, głównie poprzez wsparcie diagnostyczne wskazujące na potencjalne nieprawidłowości w pracy instalacji, przejawiające się nieprawidłowymi zużyciami energii lub czasami wysterowań urządzeń wykonawczych.

Z tego powodu trzecia z tez dodatkowych sformułowana została w postaci:

Zaproponowane podejście do badania efektywności i jakości procesów produkcyjnych, bazujące na agregacji danych i automatycznym tworzeniu profili umożliwia wykrywanie anomalii w pracy instalacji przemysłowych w produkcji dyskretnej.

Do przeprowadzenia badań i udowodnienia prawdziwości przedstawionych tez sformułowano dodatkowo następujący zestaw założeń i ograniczeń:

- Analizowane dane pochodzić muszą z instalacji przemysłowej pracującej w produkcji dyskretnej, w której rejestracja sygnałów z urządzeń wykonawczych podłączonych do modułów wejścia i wyjścia prowadzona jest przez system sterowania (np. PLC) lub akwizycji danych (np. SCADA),
- Rozpatrywana instalacja produkcyjna pracować może z wykorzystaniem dowolnego medium zasilającego (np. sprężonego powietrza, energii elektrycznej), a pomiar zużycia tego medium musi być realizowany w jednym punkcie obejmującym wszystkie odbiorniki tej energii (wskazując sumaryczne zużycie dla całej instalacji),
- Rozpatrywana instalacja musi być wyposażona w interfejsy do danych procesowych (np. kontrolery procesu), umożliwiające wystawianie danych procesowych o aktywności pojedynczych urządzeń wykonawczych, stempla lub dowolnego innego znacznika czasowego oraz pochłanianej energii w postaci zużycia całkowitego (przyrostowego) lub chwilowego.

1.4. Opis ogólny prowadzonych badań

Celem badań było opracowanie zestawu metod do analizy danych produkcyjnych, których rezultaty mogą zostać wykorzystane do poprawy efektywności i jakości pracy stanowisk produkcyjnych w wielowariantowej produkcji dyskretnej. Rozwijane metody bazowały na technikach klasteryzacji i statystyki oraz uwzględniały automatyczne rozpoznawanie wytwarzanych wariantów produktów. Prace badawcze skoncentrowane były wokół zagadnień:

- Przeglądu istniejących metod, technik i podejść stosowanych do analizy danych produkcyjnych, służących do wykrywania wariantów produkcji, wspierających

wykrywanie nieprawidłowości w pracy instalacji produkcyjnych, systematyzujących podejścia do nowoczesnych systemów wytwarzania w oparciu o modele informacyjne oraz umożliwiającym powiązanie danych produkcyjnych z wiedzą technologiczną z wykorzystaniem tych modeli,

- Akwizycji danych procesowych z poziomu urządzeń wykonawczych i kontrolerów procesów, w szczególności sterowników PLC,
- Opracowania danych zagregowanych, umożliwiającym redukcję złożoności algorytmów analitycznych oraz zapewniającym uniwersalność stosowanych metod niezależnie od analizowanej instalacji przemysłowej,
- Analizy danych zagregowanych i surowych w celu rozwiązania wybranych problemów badawczych: automatycznego grupowania wariantów ze względu na ich specyfikę z wykorzystaniem różnych wersji klasteryzacji opartej o metodę *k*-Means, automatycznej detekcji i klasyfikacji wariantów produkcyjnych, automatycznego wykrywania nieprawidłowości w pracy instalacji oraz automatycznego wykrywania źródeł anomalii,
- Modelowania składowych instalacji przemysłowych z wykorzystaniem formatu Automation Markup Language (AutomationML) i standardu ANSI/ISA-95, umożliwiającym przechowywanie i wymianę informacji o technologiach,
- Wykorzystania przestrzeni adresowej interfejsu OPC UA do powiązania danych procesowych z technologią wykonania instalacji produkcyjnej,
- Wizualizacji wyników przetwarzania i udostępnienia ich do dalszego przetwarzania.

Zadania analityczne realizowane przez system opracowane zostały po uprzednio przeprowadzonych badaniach na różnych typach danych pozyskiwanych ze zbiorów testowych i z rzeczywistych instalacji produkcyjnych. W szczególności wykorzystywane były dane treningowe z instalacji eksperymentalnych (produkcyjnych, składających się z rzeczywistych kontrolerów procesu i urządzeń wykonawczych zasilanych m.in. energią elektryczną i sprężonym powietrzem) oraz dane procesowe z różnych instalacji produkcyjnych wykorzystywanych w produkcji dyskretniej, m.in. w branży produkcji samochodowej i narzędziowej.

Rezultaty z badań nad przetwarzaniem danych i możliwościami ich zastosowania przedstawione zostały w tekstach publikowanych w czasopismach i na konferencjach naukowych:

Czasopisma

- Cupek, R., Ziebinski, A., Zonenberg, D., & Drewniak, M. (2018). Determination of the machine energy consumption profiles in the mass-customised manufacturing. *International Journal of Computer Integrated Manufacturing*, 31(6), 537-561:
 - Impact Factor: 4.9
- Cupek, R., Ziebinski, A., Drewniak, M., & Fojcik, M. (2019). Knowledge integration via the fusion of the data models used in automotive production systems. *Enterprise Information Systems*, 13(7-8), 1094-1119.
 - Impact factor: 5.12
- Cupek, R., Drewniak, M., Ziebinski, A., & Fojcik, M. (2019). “Digital Twins” for highly customized electronic devices—Case study on a rework operation. *IEEE Access*, 7, 164127-164143..
 - Impact factor: 3.367

Konferencje

- Cupek, R., Ziębiński, A., Drewniak, M., & Fojcik, M. (2018, March). Improving KPI based performance analysis in discrete, multi-variant production. In *Asian Conference on Intelligent Information and Database Systems* (pp. 661-673). Springer, Cham.
 - CORE B
- Cupek, R., Drewniak, M., & Steclik, T. (2021, June). Data Preprocessing, Aggregation and Clustering for Agile Manufacturing Based on Automated Guided Vehicles. In *International Conference on Computational Science* (pp. 458-470). Springer, Cham.
 - CORE A

- Cupek, R., Duda, J., Zonenberg, D., Chłopaś, Ł., Dziędziel, G., & Drewniak, M. (2017, September). Data mining techniques for energy efficiency analysis of discrete production lines. In *International Conference on Computational Collective Intelligence* (pp. 292-301). Springer, Cham.
 - CORE C
- Cupek, R., Ziebinski, A., & Drewniak, M. (2018, September). Application of decision trees for quality management support. In *International Conference on Computational Collective Intelligence* (pp. 67-78). Springer, Cham.
 - CORE C
- Cupek, R., Ziębiński, A., Drewniak, M., & Fojcik, M. (2018, March). Estimation of the Number of Energy Consumption Profiles in the Case of Discreet Multi-variant Production. In *Asian Conference on Intelligent Information and Database Systems* (pp. 674-684). Springer, Cham.
 - CORE C
- Cupek, R., Drewniak, M., & Ziebinski, A. (2019, October). Information models for a new generation of manufacturing systems-a case study of automated guided vehicle. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 858-864). IEEE.
 - CORE B
- Cupek, R., Ziębiński, A., Drewniak, M., & Fojcik, M. (2017, September). Application of OPC UA Protocol for the Internet of Vehicles. In *International Conference on Computational Collective Intelligence* (pp. 272-281). Springer, Cham.
 - CORE C
- Cupek, R., Ziebinski, A., & Drewniak, M. (2017, March). An OPC UA server as a gateway that shares CAN network data and engineering knowledge. In *2017 IEEE International Conference on Industrial Technology (ICIT)* (pp. 1424-1429). IEEE.
 - CORE C
- Drewniak, M., Gabryś, M. (2017, September). Cognitive Maintenance and Polymorphic Production as the Leading Industry 4.0 Paradigms. In *International Conference Mechatronics* (pp. 101-110). Springer, Cham.

- Cupek, R., Drewniak, M., & Zonenberg, D. (2014, June). Online energy efficiency assessment in serial production-statistical and data mining approaches. In 2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE) (pp. 189-194). IEEE.

1.5. Zawartość pracy

Niniejsza praca zorganizowana została w następujący sposób: rozdział 2. opisuje podstawy teoretyczne zjawisk i problemów rozwiązywanych w trakcie badań, ze szczególnym wskazaniem na specyfikę stosowanych współcześnie systemów produkcyjnych, możliwości komunikacyjne, rozwiązania zdolne do przetwarzania danych w sposób umożliwiający osiągnięcie rezultatów końcowych, techniki analityczne możliwe do zastosowania w rozwiązaniu oraz stan badań nad dziedziną. W rozdziale 3. opisane zostało podejście do zastosowania współczesnych modeli danych do powiązania wiedzy technologicznej o sposobie wykonania instalacji produkcyjnej z danymi procesowymi gromadzonymi w trakcie produkcji, a na podstawie których prowadzone są dalsze analizy i rozwój metod opisanych w kolejnych rozdziałach. Rozdział 4. zawiera opis rezultatów z badań nad technikami agregacji danych procesowych i metodami automatycznego wykrywania profili produkcyjnych z wykorzystaniem wybranych technik eksploracji danych. Rozdział 5. prezentuje metodę wykrywania anomalii w pracy instalacji produkcyjnych z zastosowaniem opracowanych w ramach rozdziału 4. wzorców technologicznych. W rozdziale 6. zaprezentowane zostały wnioski z prowadzonych badań i osiągniętych rezultatów. Rozdział 7. zawiera źródła literaturowe wykorzystywane podczas badań i do stworzenia niniejszej pracy.

2. ANALIZA DZIEDZINY I PRZEGLĄD ISTNIEJĄCYCH ROZWIĄZAŃ

2.1. Systemy realizacji produkcji

Jednym z podstawowych narzędzi wykorzystywanym we współczesnym przemyśle są systemy realizacji produkcji (ang. Manufacturing Execution System, MES). Są to rozwiązania informatyczne, które w pierwotnej formie przedstawione zostały jako koncepcyjne systemy cyfrowe wspomagające wytwarzanie (ang. Computer-Integrated Manufacturing, CIM) [8], a które dzięki rozwojowi technik i usług informatycznych rozpowszechnione zostały w latach 90-tych XX wieku. Ich głównym celem jest wsparcie monitorowania oraz zarządzanie realizacją wytwarzania w zakładach produkcyjnych [9]. Zadania te obejmują wszystkie etapy procesów produkcyjnych, w których materiały surowcowe lub komponenty służą do wytworzenia produktów końcowych [10]. W miarę rozwoju systemów MES pojawiało się coraz więcej ich funkcjonalności, a obecnie wyróżnia się 11 głównych funkcji rozwiązań tej klasy:

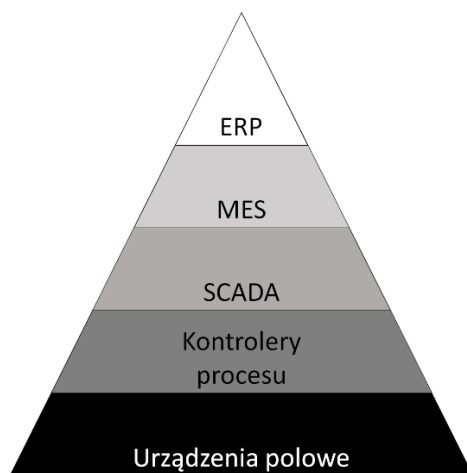
- Alokowanie i śledzenie statusów zasobów produkcyjnych,
- Szeregowanie zadań produkcyjnych,
- Dysponowanie jednostkami produkcyjnymi,
- Kontrola dokumentów,
- Zbieranie i gromadzenie danych produkcyjnych,
- Zarządzanie personelem produkcyjnym,
- Zarządzanie jakością,
- Zarządzanie utrzymaniem infrastruktury produkcyjnej,
- Zarządzanie procesami produkcyjnymi,
- Śledzenie i genealogia produktów,
- Analiza wydajności produkcji [11]

Dla potrzeb niniejszej pracy system realizacji produkcji rozumiany będzie jako system informatyczny, realizujący zadania monitorowania danych procesowych z układów wykonawczych, analizujący te dane dla potrzeb rozwiązywania wybranych problemów badawczych, w szczególności oceny jakości pracy stanowisk wykonawczych i wykrywania anomalii w ich pracy oraz umożliwiający powiązanie rezultatów z tych analiz z wiedzą o technologii wykonania instalacji produkcyjnej.

Systemy realizacji produkcji i środowiska przemysłowe przeżywają obecnie fazę istotnych przemian. Dotychczasowa dyskretna produkcja masowa (ang. mass production), w której wytwarzanie realizowane było przez zamknięte i odpowiednio do tego przygotowane ciągi technologiczne, wytwarzające seryjnie produkty w znanej i zwykle istotnie ograniczonej liczbie wariantów, stopniowo zastępowana jest produkcją masowo-dostosowaną (ang. mass-customized), w której procesy produkcyjne muszą być prowadzone elastycznie, zapewniając wytwarzanie wielowariantowe, a często także krótko- lub nawet jednoseryjne. Wynika to przede wszystkim ze zmian rynkowych, dających możliwość klientom końcowym swobodnej konfiguracji składanego zamówienia oraz rozpowszechnieniu usług i rozwiązań informatycznych wspierających to podejście [12]. Aby osiągnąć tę elastyczność konieczne jest zastosowanie różnorodnych technik z obszaru informatyki, automatyki i elektroniki przemysłowej, które nieustannie muszą być wdrażane przez wytwórców. Należą do nich między innymi przemysłowy internet rzeczy (ang. Industrial Internet of Things, IIoT), integracja różnego rodzaju maszyn i systemów produkcyjnych, analiza danych procesowych czy tworzenie wirtualnego bliźniaka wytwarzanego produktu lub instalacji produkcyjnej (ang. Digital Twin) [13]. Niejednokrotnie rodzą one jednak poważne problemy integracyjne, w których obecnie stosowane rozwiązania przemysłowe, np. systemy MES lub układy automatyki nie są dostosowane do ich wspierania. Z tego powodu transformacja z produkcji masowej do masowo-dostosowanej wpływa nie tylko na samą sferę produkcyjną, ale zmienia również zasady związane z wymianą informacji pomiędzy systemami przemysłowymi. Wymagana jest ona do stosownej współpracy różnego typu rozwiązań informatycznych, analizy danych procesowych czy wsparcia budowania historii produkcji (ang. traceability) [14].

Wieloletnie podejście do produkcji przez seryjność i niską wariantowość produktów końcowych utrwaliło stosowanie jednego z najczęściej wykorzystywanych modeli produkcyjnych, tj. klasycznej piramidy automatyzacji. Architektura ta zakłada pionową integrację usług pomiędzy poszczególnymi systemami przemysłowymi dla potrzeb wymiany danych od systemu zarządzania majątkiem przedsiębiorstwa (ang. Enterprise Resource Planning, ERP) po urządzenia wykonawcze. Dane wymieniane pomiędzy warstwami ograniczone były do niezbędnych i koniecznych dla pracy odpowiednich systemów, a stosowanie tych warstw zależne było od faktycznie występującej infrastruktury na terenie obiektu przemysłowego. Architektura ta wprowadziła uproszczenie podziału systemów informatycznych stosowanych w przemyśle i ich

rozgraniczenie w stosowaniu w poszczególnych warstwach, np. MES w warstwie realizacji produkcji czy sterowników swobodnie programowalnych PLC w warstwie kontrolerów procesu [15].

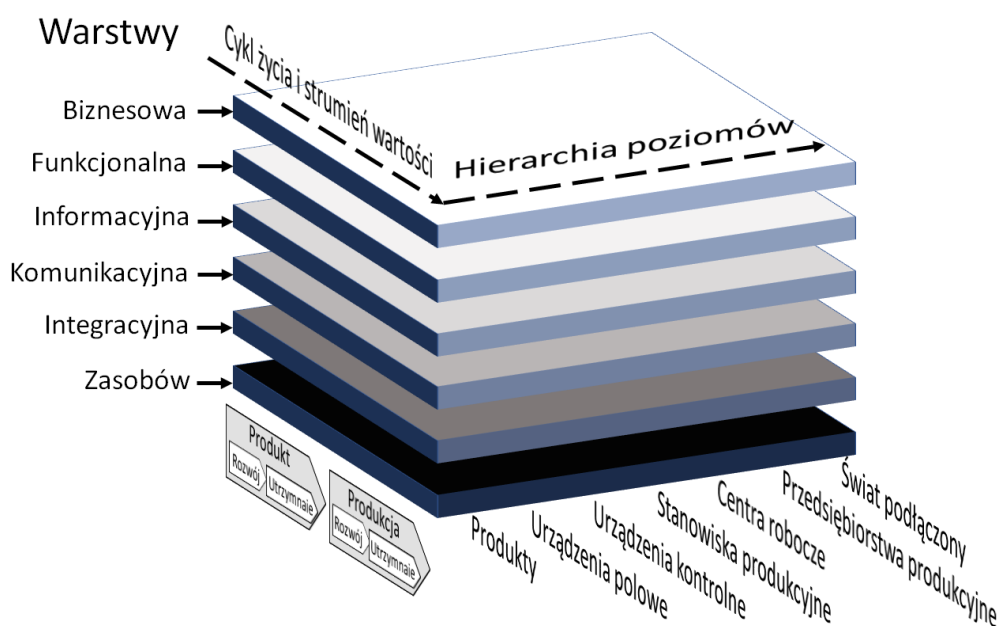


Rysunek 2.1.1. – piramida automatyzacji

Zmieniające się wymagania dla współczesnych systemów przemysłowych powodują, że model klasycznej piramidy nie zawsze spełnia wymagania stawiane przed elastycznym wytwarzaniem. Wykorzystywanie różnych informacji i różnorodnych standardów wymiany danych powoduje ogromne trudności lub niejednokrotnie uniemożliwia integrację nowoczesnych systemów produkcyjnych, np. systemów MES, cyfrowego wsparcia utrzymania CMMS, ERP czy samych systemów automatyki zakładowej. Dodatkowo do puli tych rozwiązań dołączają coraz częściej systemy wspomagające projektowanie, np. CAD (ang. Computer-Aided Design), zarządzanie cyklem życia produktu PLM (ang. Product Lifecycle Management) czy narzędzia symulacyjne, dostarczające dane o zasobach projektowych lub wiedzy technologicznej. Jednocześnie, przez odzwierciedlenie statycznej hierarchii usług produkcyjnych, np. ERP ponad MES, model piramidy znacząco ogranicza elastyczność w nowoczesnych systemach produkcji i podnosi koszty integracji pomiędzy tymi systemami, głównie przez pionową wymianę danych pomiędzy wybranymi warstwami i zachowanie odpowiednich połączeń. Wreszcie, co z punktu widzenia wymiany danych jest najważniejsze, architektura ta zakłada jedynie przepływ informacji na etapach realizacji produkcji, bez budowania wiedzy o wytwarzanej produkcie, stosowanych technologiach lub bez modelowania przepływu informacji pomiędzy warstwami. W dobie produkcji elastycznej i krótkoseryjnej zapewnienie odpowiedniej integracji danych pomiędzy systemami informatycznymi nie

może sprowadzać się jedynie do ujednolicenia formatów danych lub odpowiedniej ich translacji, ale również do opracowania odpowiednich mechanizmów ich wymiany, tak, aby związać dane procesowe (wartości bieżące) z danymi technologicznymi (wiedzą o procesie, urządzeniach czy ich utrzymaniu). Ponieważ Przemysł 4.0 budowany jest w oparciu o nowoczesne rozwiązania informatyczne, w tym systemy zorientowane usługowo (ang. Service-Oriented Architecture, SOA), podejście do elastycznego i adaptacyjnego wytwarzania wymaga zaprojektowania zgodnego z tą właśnie filozofią [16].

Odpowiedzią na zapotrzebowanie na elastyczne systemy produkcyjne jest uwspólnienie podejścia do wymiany danych z wykorzystaniem modeli informacyjnych nowej generacji. Jednym z nich jest RAMI 4.0 - Referencyjny Model Architektury dla Przemysłu 4.0 (ang. Reference Architecture Model for Industry 4.0), zaproponowany przez niemiecki instytut badawczy Plattform Industrie-4.0, zajmujący się opisywaniem elastycznych technik wytwarzania przemysłowego. Strukturyzuje on istniejące standardy wymiany danych pomiędzy składowymi systemów produkcyjnych, wyszukuje brakujące połączenia i wskazuje na obszary wymagające standaryzacji. Dodatkowo systematyzuje on podejścia do stosowania koncepcji Przemysłu 4.0 w praktyce. Trójwymiarowa prezentacja modelu RAMI4.0 przedstawiona została na rysunku 2.1.2.



Rysunek 2.1.2. – architektura RAMI4.0

Oś pionowa architektury RAMI4.0 zorganizowana jest w sześciu warstwach, skupiających się na różnych aspektach serwisów produkcyjnych. Modeluje ona przepływ informacji od warstwy fizycznej zasobów po wykorzystanie wiedzy na ich temat na poziomie biznesowym. Pierwszą z warstw jest warstwa zasobów (ang. Asset layer), reprezentująca zasoby wymagane do produkcji, a które będą wirtualizowane w wyższych warstwach będą wirtualizowane. Są to: infrastruktura produkcyjna składająca się z urządzeń wykonawczych i systemów kontrolnych, zasoby fizyczne wymagane do produkcji, technologia wraz z jej organizacją opisana za pomocą dokumentów, oprogramowania i nastaw procesowych oraz personel techniczny wymagany w procesach produkcyjnych. Drugą warstwę stanowi integracja (ang. Integration layer). Odpowiada ona przejściu ze świata rzeczywistego do świata informacji i opisuje zasoby z pierwszej warstwy, które są wymagane do zaimplementowania funkcji ich opisujących lub wykorzystujących. Warstwa ta zapewnia więc reprezentację odpowiednich zasobów poprzez informację o nich. Dane o zasobach mogą być gromadzone w postaci odczytów z urządzeń fizycznych, np. czujników, systemów identyfikacji optycznej czy RFID, konwerterów sygnałów, systemów kontrolnych i generujących zdarzenia na podstawie fizycznych właściwości zasobów, jak również mogą być wprowadzane przez człowieka. Warstwa komunikacyjna (ang. Communication layer) definiuje dostęp do informacji i funkcji tak, aby zapewnić wymianę danych pomiędzy poszczególnymi zasobami z wykorzystaniem technologii i protokołów komunikacyjnych bazujących na podejściu zorientowanym usługowo (np. SOA). Opisuje dodatkowo, jakiego typu dane są używane i udostępniane dla bieżących potrzeb wymiany informacji. Warstwa informacyjna (ang. Information layer) definiuje operacje na danych wykorzystywanych podczas działania na zasobach, od mechanizmów gwarantujących integralność i spójność, przez definicję modeli formalnych i zasad przetwarzania, po wykonywanie zasad. Dzięki tym operacjom możliwe jest przekształcenie surowych danych w informacje i wiedzę. Dodatkowo, zapewnia przygotowanie danych do wymiany z kolejnymi warstwami, w szczególności poprzez transformację do odpowiednich form i formatów oraz przekazaniu ich za pomocą przygotowanych interfejsów zorientowanych usługowo. Warstwa funkcjonalna (ang. Functional layer) opisuje funkcje zasobów wykorzystywanych w procesie. Zapewnia formalne opisy tych funkcji, modeluje środowisko dla serwisów przemysłowych i biznesowych, stanowi warstwę uruchomieniową dla tych funkcji i integruje je z odpowiednimi zasobami. Ostatnia z warstw, biznesowa (ang. Business Layer), definiuje komercyjne aspekty usług produkcyjnych i opisuje m.in. zlecenie zadań produkcyjnych,

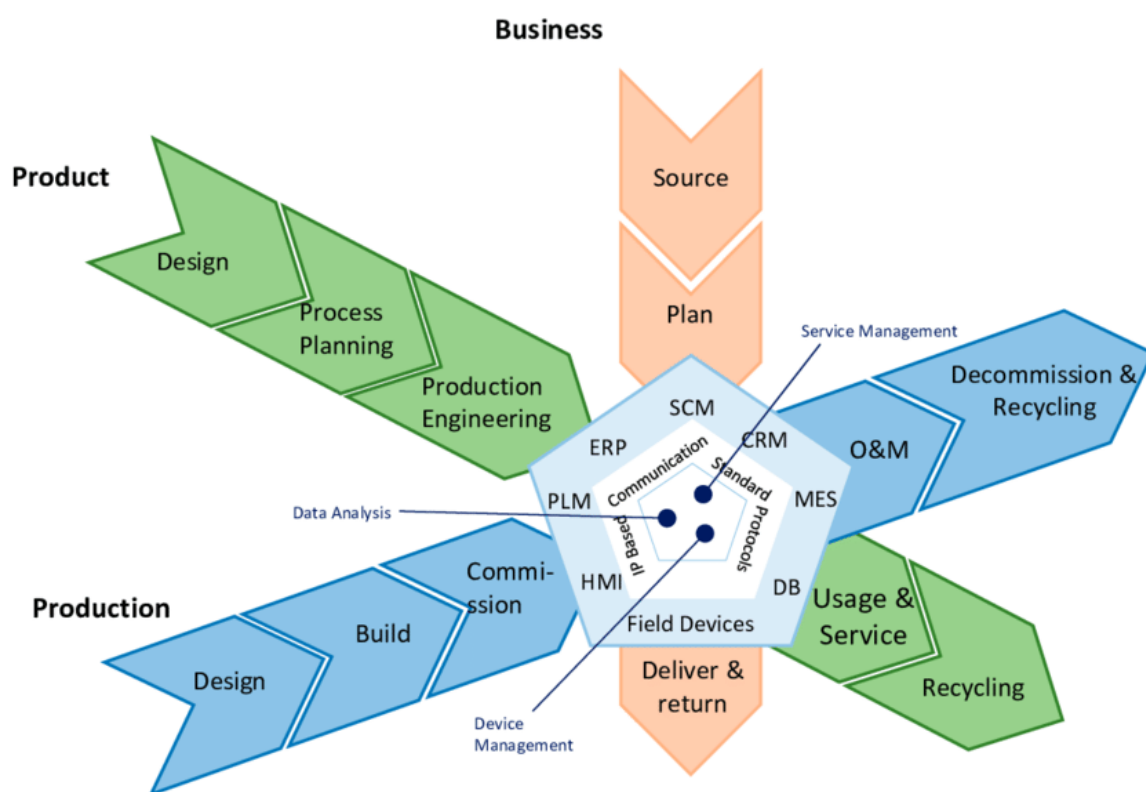
zaopatrzenie, koszty, dostępność zasobów i integralność łańcucha wartości dodawanych (ang. Value-added chain), od materiałów surowcowych po gotowy produkt.

Pierwsza z osi poziomych reprezentuje cykl życia produktu i strumień wartości z nim związany, skupiając się na projektowaniu, produkcji i utrzymaniu wyrobu gotowego, jak również projektowaniu, wdrożeniu i utrzymaniu procesu produkcyjnego. Uwzględnia ona etapy rozwoju produktu i produkcji z wykorzystaniem technik wspomagających takich jak symulacje, etap produkcji z pełną kontrolą parametrów procesowych i śledzeniem danych produkcyjnych oraz etapy utrzymania produktu i procesu, od przepływu zasobów związanych z oprogramowaniem czy dokumentami po warunki użytkowania, serwisowania czy utylizacji. Druga z osi obrazuje funkcjonalną hierarchię poziomów wytwarzania, od produktu, przez infrastrukturę produkcyjną, po świat podłączony z wytwarzanymi wyrobami czy procesem produkcyjnym. Odzwierciedla ona rzeczywistą strukturę produkcyjną dostępną na terenie obiektów przemysłowych, zależną od warunków lokalnych.

Choć model RAMI4.0 jest ogólnym modelem prezentującym holistyczne podejście do nowoczesnych systemów realizacji produkcji zgodnych z paradygmatem Przemysłu 4.0., wdrażanie go w kompletnej postaci zwykle jest niemożliwe z uwagi na już istniejącą infrastrukturę i systemy produkcyjne, pracujące w filozofii produkcji masowej lub dedykowanej. Z tego powodu stosowanie modelu RAMI4.0 polega raczej na jego dostosowywaniu do warunków i technologii produkcyjnych, w których ma być uruchomiony, a samo wdrażanie wymaga modernizacji istniejących rozwiązań lub uzupełnianiu ich o nowe fragmenty tak, aby osiągać docelowy model jak najbardziej zbliżony do referencyjnego.

Oprócz modelu RAMI4.0, opracowane zostały inne referencyjne modele wspierające nowoczesne systemy wytwarzania. Pierwszym z nich jest Smart Manufacturing Ecosystem (SME) opracowany przez National Institute of Standards and Technology w Stanach Zjednoczonych. Model ten skupia się na trzech wymiarach odpowiadających różnym aspektom produkcyjnym: wymiarze produkcji wraz z cyklem jej życia od projektowania przez budowę, uruchomienie i wykorzystanie po demontaż i utylizację, wymiarze produktu z podobnym przepływem wartości, uzupełnionym o kroki użytkowania i serwisowania oraz wymiarze biznesowym, definiującym obieg informacji od składania zamówień, przez planowanie produkcji na poziomie systemów ERP po

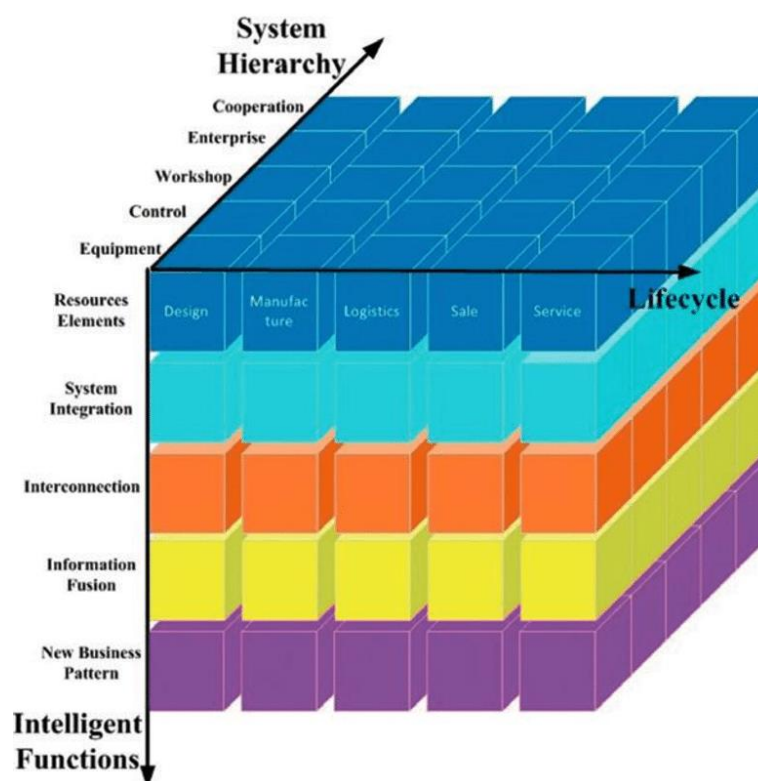
wypuszczenie zamówień na rynek końcowy. Centralną częścią tej architektury jest klasyczna piramida automatyzacji, w której przepływ wertykalny zapewniony jest przez różne interfejsy i technologie, m.in. systemy cyberfizyczne czy technologie łączności bezprzewodowej. W ramach tej części zdefiniowane zostały standardy, interfejsy i protokoły komunikacyjne wykorzystywane pomiędzy poszczególnymi składowymi systemów produkcyjnych, od urządzeń wykonawczych po ERP. Dane z poszczególnych wymiarów mogą być przekazywane pomiędzy nimi tak, aby wspierać transfer wiedzy i obieg dokumentów, np. wyniki symulacji produktowych z etapu planowania wpływać mogą na procesy przygotowywania produkcji, a wnioski z użytkowania i serwisowania produktów końcowych mogą uzupełniać kroki projektowania produktów. Poszczególne wymiary i przepływ danych w architekturze SME przedstawione zostały na rysunku 2.1.3.



Rysunek 2.1.3. – architektura Smart Manufacturing Ecosystem [17]

Innym przykładem architektury jest Smart Manufacturing Standardization (SMS) przygotowany przez chiński Intelligent Manufacturing System Association (IMSA). Model ten jest zbliżony do RAMI4.0 poprzez definicję podobnych poziomów hierarchii systemów produkcyjnych i elementów związanych z cyklem życia produktu i instalacji, różni się jednak definicją funkcjonalności, wyrażonych w RAMI4.0 jako warstwy. SMS opisuje integrację i fuzję informacji z wykorzystaniem aspektów bezpieczeństwa

cybernetycznego, technik CPS i IoT czy technik Big Data. Schemat architektury SMS przedstawiony został na rysunku 2.1.4.



Rysunek 2.1.4. – architektura Smart Manufacturing Standardization [18]

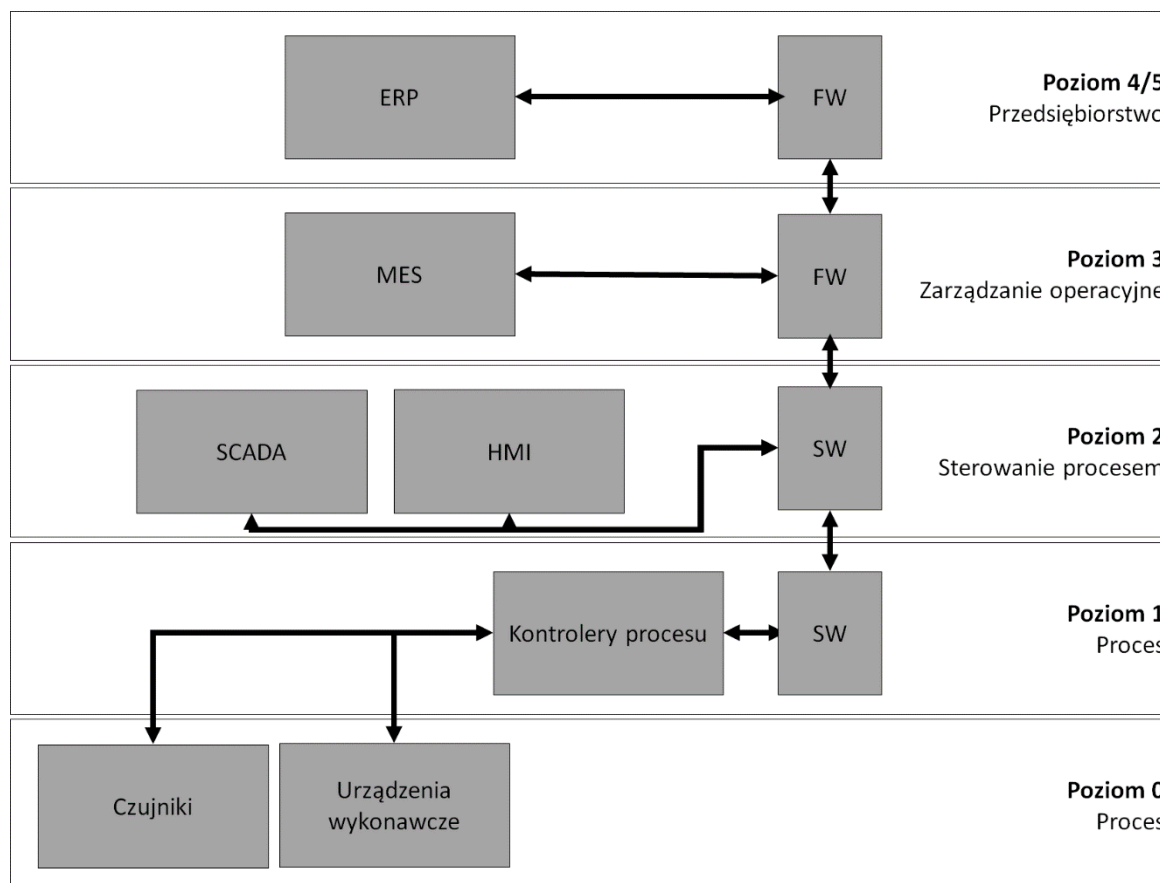
Przedstawione architektury różnią się od siebie w pewnych aspektach, np. model SME funkcjonalnie definiuje te same lub podobne zagadnienia związane z cyklem życia produktu i instalacji czy warstwy co RAMI4.0, lecz różni się definicjami możliwości łączenia i tworzenia interfejsów, wskazując np. na bezpieczny dostęp do danych procesowych z wykorzystaniem sieci bezprzewodowych i technologii nienaruszalności danych blockchain. Z kolei model IMSA opisuje podobne poziomy hierarchii co RAMI4.0, jednak znacząco inaczej definiuje strumień wartości i cykl życia produktu, obejmując np. sprzedaż na rynku odbiorów, lecz nie ujmując procesów utrzymania wytwarzania [19] [20]. Ponieważ model RAMI4.0 naturalnie wyewoluował z hierarchii poziomów wytwarzania i strumienia wartości oraz cykli życia produktu i procesu, co pokrywa się z założeniami niniejszej rozprawy, w której dane z procesu produkcyjnego, po przekształceniu w informacje oraz wiedzę mogą wspierać utrzymanie tego procesu, w dalszej części pracy zdecydowano o wykorzystaniu tej właśnie architektury do wpasowania rozwijanych metod w nowoczesnym systemie realizacji produkcji.

Oprócz zmian w systemach realizacji produkcji, które muszą wspierać elastyczne wytwarzanie, a które mogą zostać zbudowane w oparciu o modele informacyjne nowej generacji, istotne przemiany zachodzą również w produkcyjnych systemach pomocniczych. Jednym z najbardziej dynamicznie rozwijających się trendów przemysłowych jest analiza danych produkcyjnych i danych z systemów zarządzania produkcją w celu wykrywania różnego rodzaju zjawisk i wzorców, wykorzystywanych następnie do budowania wiedzy o procesach produkcyjnych i ciągłej poprawy efektywności, wydajności i jakości produkcji. Przykładem takiego trendu jest predykcyjne i prewencyjne utrzymanie instalacji (ang. predictive and preventive maintenance), mające na celu obserwację bieżącej pracy urządzenia wykonawczego lub stanowiska produkcyjnego, ocenę jakości i stabilności tej pracy w odniesieniu do wykorzystywanej technologii i wykrywanie anomalii [21]. Anomalie te mogą z jednej strony przejawiać się chwilowymi zaburzeniami w pracy, widocznymi w postaci nieprawidłowych wartości w danych procesowych, a z drugiej jako postępująca degradacja, mogąca prowadzić do poważnej awarii, a w rezultacie zatrzymania pracy i konieczności podjęcia kosztownych napraw. Ulepszanie diagnostyki instalacji produkcyjnych jest pożądanym podejściem przez zakłady produkcyjne, prowadzącym nie tylko do zapobiegania poważnym awariom ale także do wydłużania czasu eksploatacji przez dostarczanie sugestii związanych z serwisowaniem, przeglądami okresowymi lub przewidywanymi momentami wymiany komponentów. Aby skutecznie wdrożyć mechanizmy prewencyjnego utrzymania instalacji konieczne jest jednak zastosowanie technik analitycznych zdolnych do przetworzenia danych produkcyjnych i wykrycia niejednokrotnie skomplikowanych zależności wiążących dane. Z tego powodu chętnie stosowane są tu metody z obszaru eksploracji danych oraz uczenia maszynowego. Niestety, ze względu na niejednokrotnie unikatowy charakter badanych zjawisk i analizowanych danych brak jest uniwersalnych narzędzi zdolnych do rozwiązywania ogólnych problemów wykrywania anomalii i degradacji pracy instalacji produkcyjnych, a dostępne rozwiązania powstają jako dedykowane do konkretnych problemów lub technologii. Te właśnie powody oraz styczność ze środowiskami produkcyjnymi w pracy zawodowej skłoniły autora niniejszej rozprawy do rozwoju metod automatycznego wykrywania błędów w pracy instalacji produkcyjnych, w szczególności realizujących zadania produkcji dyskretnej i wielowariantowej.

2.2. Modelowanie usług i danych produkcyjnych z wykorzystaniem podejścia obiektowego

Rozwój i rozpowszechnienie systemów realizacji produkcji, stosowanie nowych technologii z zakresu komunikacji, przetwarzania i wykorzystania danych oraz zmiana w filozofii wytwarzania powodują, że architektura i sposób wymiany danych pomiędzy usługami przemysłowymi wymagają podejścia innego, niż stosowane było dotychczas, np. w najpowszechniej stosowanym podejściu opartym o piramidę automatyzacji. Choć potrzeba ta jest coraz silniej widoczna w różnych branżach przemysłowych, jej ewolucja do obecnej postaci stanowiła proces rozwijany przez wiele lat, a jej efektem były różne koncepcje i sposoby modelowania, m.in. hierarchii poziomów wytwarzania, usług czy komunikacji pomiędzy nimi.

Pierwszym z problemów zaobserwowanych w przemyśle wraz z pojawieniem się różnego rodzaju systemów informatycznych była konieczność wymiany danych pomiędzy nimi. Istotne różnice w funkcjonalności systemów wymagały wymiany danych o różnej specyfice, np. sygnały procesowe z urzędzeń wykonawczych wykorzystywane były przez kontrolery procesu, a dane o zleceniach produkcyjnych do wykonania wysyłane przez systemy ERP przetwarzane były przez systemy realizacji produkcji MES. Jednocześnie informacje wymieniane pomiędzy systemami z reguły były bezużyteczne lub nadmiarowe w przypadku połączeń pomiędzy niezwiązanymi ze sobą logicznie komponentami, np. dane o zleceniach były niepotrzebne dla działalności operacyjnej kontrolerów procesów, wymagających zamiast tego szczegółowych nastaw do pracy. Drugim istotnym aspektem pojawiającym się wraz z rozwojem Internetu i dostępnością różnego rodzaju sieci i urzędzeń sieciowych były problemy bezpieczeństwa danych i dostępu do nich. Konieczne stało się oddzielenie procesu i związanych z nim krytycznych ze względów bezpieczeństwa i ciągłości produkcji danych od klasycznych sieci biurowych oraz oddzielenie danych związanych z zarządzaniem produkcją od innych danych firmowych. Wobec powyższych, w latach 90. zespół badaczy z Uniwersytetu Purdue opracował referencyjną architekturę przedsiębiorstwa (ang. Purdue Enterprise Reference Architecture, PERA), nazwaną modelem Purdue. Opisywał on różne poziomy infrastruktury przedsiębiorstwa wraz ze ich sposobami zabezpieczania dostępu do sieci [22]. Uproszczony schemat modelu Purdue przedstawiony został na rysunku 2.2.1.



Rysunek 2.2.1. – uproszczony schemat modelu Purdue

Poziom 0. składał się z urządzeń wykonawczych procesu, sterowanych i komunikujących się z lokalnymi kontrolerami procesu z poziomu 1. Komunikacja ta zrealizowana była w oparciu o różnorakie interfejsy i protokoły przemysłowe oraz realizowana była bez udziału współcześnie rozumianych sieci IT. Jednocześnie poziom 1. zakładał możliwość rozproszenia systemów sterowania, a więc współdzielenie zasobów obliczeniowych i pamięci kontrolerów procesu, z tego powodu konieczne było przewidzenie wspólnej sieci komputerowej, reprezentowanej przez odpowiednie urządzenia sieciowe (na rysunku blok SW odpowiadający urządzeniu klasy switch). Poziom 2. stanowił sterowanie procesem z wykorzystaniem systemów SCADA i HMI (ang. Human-Machine Interface, HMI), zawierał więc rozwiązania bazodanowe, elementy GUI (ang. General User Interface) oparte o WWW (ang. World Wide Web) oraz zdalne jednostki terminalowe RTU (ang. Remote Terminal Units). Funkcjonalne oddzielenie tej części od procesu podyktowane było bezpieczeństwem jego ciągłości, stąd wymiana danych pomiędzy poziomem 1. odbywała się z wykorzystaniem separujących sieć urządzeń (na rysunku blok SW). Poziom 3. odpowiadał zarządzaniu produkcją i śledzeniu

jej parametrów. Ponieważ systemy klasy MES należące do tego poziomu były klasycznymi aplikacjami komputerowymi i stanowiły najczęściej indywidualne rozwiązania od różnych producentów, a jednocześnie były krytyczne dla ciągłości i prawidłowości produkcji ze względu na swoją specyfikę, wymiana danych z poziomem 2. wymagała silnych zabezpieczeń sieciowych (reprezentowanych na rysunku przez blok FW odpowiadający zabezpieczeniu firewall). Najwyższymi poziomami, 4. i 5., były warstwy odpowiadające zarządzaniu przedsiębiorstwem z wykorzystaniem systemów ERP, oddzielone od świata zewnętrznego odpowiednimi zabezpieczeniami [23].

Model Purdue stał się wzorcem do implementacji komunikacji i usług przemysłowych pomiędzy poziomami wytwarzania i z powodzeniem stosowany był w przemyśle przez wiele lat. Choć architektura Purdue stanowiła ikoniczny standard w systemach przemysłowych, opisywany przez nią przepływ danych nie może być wprost wykorzystywany we współczesnych środowiskach produkcyjnych, głównie za sprawą niejednoznacznego podziału na warstwy hierarchii infrastruktury i zdolności komunikacyjne poszczególnych urządzeń przemysłowych. Nowoczesne systemy pomiarowe i urządzenia wykonawcze mogą być wyposażone we własną inteligencję oraz bezpośrednie połączenie z warstwami wyższego poziomu, nie ma więc potrzeby zachowywania wertykalnej komunikacji od poziomu 0 do poziomu 5. Dodatkowo, konieczność adaptacyjnego dostosowywania produkcji do potrzeb rynkowych spowodowały, że konieczne było opracowanie podejścia zorientowanego na usługi i dynamicznie dostosowującego się do bieżących warunków produkcyjnych [24]. Wreszcie, rozwój technik nowoczesnego wsparcia produkcji, takiego jak cyfrowe wsparcie zarządzania utrzymaniem instalacji CMMS czy wsparcie projektowania CAD spowodowały konieczność przedefiniowania strumieni i przekazywanych danych, którymi odtąd mogły być również zasoby multimedialne, informacje o technologii oraz wiedza ekspercka. Ponieważ celami niniejszej pracy było wykorzystanie danych procesowych z produkcji wielowariantowej i krótkoseryjnej do budowania informacji diagnostycznych o produkcji, procesie i do wsparcia utrzymania instalacji produkcyjnej, zbudowanie rozwiązania informatycznego o model Purdue nie zapewniłoby tej możliwości.

Informacje o anomaliach w pracy stanowisk produkcyjnych mogą zasilać wiedzę i usprawniać pracę personelu zakładowego, np. lokalnych służb utrzymania ruchu lub inżynierii produkcyjnej. Przekazywanie rezultatów z działania algorytmów analitycznych

odbywać się może zarówno bezpośrednio do personelu za pomocą klasycznych technik stosowanych w interfejsach maszyna-człowiek, np. widoków www czy ekranów operatorskich, jak również systemowo, poprzez propagację wyliczanych wartości do systemów wyższego poziomu, np. MES lub ERP. Z pomocą mogą przyjść tu modele informacyjne wspierające elastyczne wytwarzanie, zgodne z architekturą RAMI4.0, w której to informacje z cyklu produkcji zasilać mogą wiedzę zarówno o docelowym produkcie jak i procesach, w których powstaje. Modelem uznawanym za jeden z kluczowych przy wdrażaniu architektury RAMI 4.0 w przemyśle jest model ANSI/ISA-95.

Model ten, jako zestaw norm opracowany został przez organizację International Society of Automation i jest standardem integrującym systemy kontrolne i zarządzania przedsiębiorstwem. Jego celem jest zamodelowanie oraz specyfikacja informacji, które mogą być wymieniane pomiędzy różnymi procesami, w szczególności produkcyjnymi i biznesowymi. Został stworzony, aby zapewnić interoperacyjność pomiędzy różnymi systemami komputerowymi wykorzystywanymi przez przedsiębiorstwa z jednoczesnym rozgraniczeniem odpowiedzialności i funkcji, które pełnią każde z nich. W ramach modelu zdefiniowane są przepływy danych, serwisy produkcyjne oraz funkcje systemów realizacji produkcji. Model bazuje na hierarchii poziomów z modelu Purdue i skupia się na połączeniach pomiędzy pięcioma warstwami systemów produkcyjnych:

- Poziomem 0, obejmującym proces produkcyjny realizowany przez urządzenia wykonawcze i monitorowany przez czujniki i systemy monitorujące,
- Poziomem 1, odpowiadającym za sterowanie procesem produkcyjnym, składającym się z systemów kontrolnych,
- Poziomem 2, monitorującym proces, wspierającym zarządzanie i sterowanie procesem, m.in. poprzez akwizycję danych procesowych i translację zleceń i planów produkcyjnych z poziomów wyższych, obejmującym systemy SCADA (ang. Supervisory Control and Data Acquisition),
- Poziomem 3, zarządzającym operacjami produkcyjnymi, składającym się z systemów MES tłumaczących zlecenia biznesowe na zadania produkcyjne,

- Poziomem 4, odpowiadającym harmonogramowaniu produkcji, planowaniu dostaw, zarządzaniu zasobami i personelem produkcyjnym, obejmującym systemy ERP.

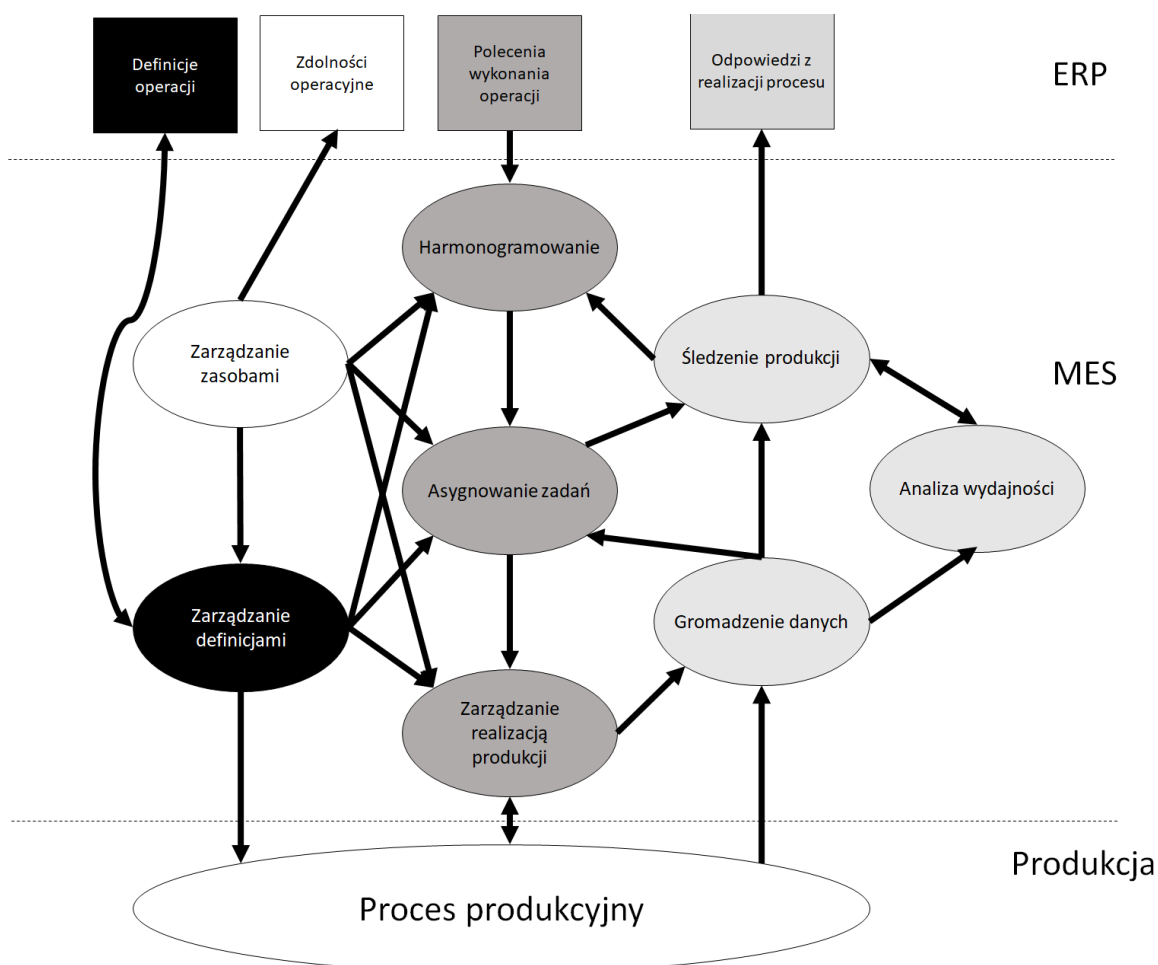
Wymiana informacji pomiędzy poziomami i systemami realizowana jest dla czterech głównych celów:

- Definicji produkcji. Definicje są informacjami opisującymi proces technologiczny jaki musi zostać wykonany w celu wytworzenia produktu końcowego. Obejmują one zasady dotyczące produkcji, opis materiałów wykorzystywanych podczas procesu, opis poszczególnych kroków produkcyjnych i ich kolejności, strukturę i właściwości infrastruktury wykorzystywanej podczas produkcji oraz wszystkie inne aspekty, które są znaczące dla procesu.
- Określania zdolności produkcyjnych. Zdolności produkcyjne kumulują informacje o wszelkiego typu zasobach i możliwościach, które w danym momencie lub ogólnie umożliwiają produkcję odpowiednich wariantów produktów. Obejmują środki takie jak maszyny i narzędzia, personel techniczny, materiały i zasoby oraz energie niezbędne do produkcji. Zdolność produkcyjna rozgraniczana jest poprzez określenie zdolności wymaganej, definiującej zaplanowane materiały i zasoby niezbędne do produkcji w zaplanowanym oknie czasowym lub wsadzie produkcyjnym, zdolności dostępnej, wskazującej na teoretyczną zdolność maksymalną chwilową lub ogólną oraz zdolności nieosiągalnej, wskazującej na stopień niewykorzystania materiałów i zasobów w procesie produkcyjnym.
- Harmonogramowania produkcji. Harmonogramy opisują zaplanowane czynności związane z realizacją produkcji. Informacje o kolejności i zadaniach produkcyjnych dekomponowane są od poziomu biznesowego, definiującego partie lub typy i wolumeny produktów końcowych do poziomu wykonawczego automatyki, w którym informacje o procesie do zrealizowania przekazywane są do urządzeń wykonawczych w postaci nastaw, ilości wykorzystywanych materiałów i zasobów czy planów wykorzystania personelu technicznego. Harmonogramowanie produkcji uwzględnia informacje o definicjach i zdolnościach produkcyjnych.

- Określania wydajności produkcji. Wydajność opisuje, co i w jakich ilościach zostało wyprodukowane. Informacje te wytwarzane są jako odpowiedź na harmonogramowanie i obejmują stopień zużycia materiałów i asygnację zasobów ludzkich do poszczególnych procesów, śledzenie nastaw i wyników pracy układów wykonawczych, monitorowanie energii oraz wszystkich innych parametrów procesowych mogących świadczyć o jakości i efektywności prowadzonych operacji. Informacje o wydajności wykorzystywane są do budowania bieżących informacji o procesie oraz gromadzone dla potrzeb historii produkcyjnych, m.in. wymaganych przez klientów końcowych.

Poszczególne modele wchodzące w skład ANSI/ISA95 opisują hierarchię infrastruktury produkcyjnej, zasoby produkcyjne, wytwarzane dobra, zlecenia i aktywności produkcyjne, harmonogramy i listy zapotrzebowań, aktywności związane z kontrolą jakości i utrzymaniem ruchu, gromadzenie i przechowywanie historycznych danych produkcyjnych oraz przepływ danych pomiędzy opisanymi poziomami dla wyszczególnionych celów. Wszystko to uwzględnia ponadto różnego rodzaju atrybuty. Opis tych modeli przygotowany został w sposób zorientowany obiektowo. Fakt ten oraz ogólność modeli powodują, że ANSI/ISA95 jest uniwersalnym standardem, możliwym do zastosowania w różnych zakresach (stopniach wykorzystania), dziedzinie (branży) oraz typu produkcji, niezależnie od tego czy ma ona charakter dyskretny, ciągły czy wsadowy.

Jednym z najważniejszych fragmentów standardu ISA-95 jest część III, opisująca realizację i utrzymanie produkcji. Zbudowana została w oparciu o typowe serwisy wykorzystywane w systemach klasy MES, a ich wyszczególnienie oraz powiązanie przedstawione zostało na rysunku 2.2.2.



Rysunek 2.2.2. – najczęściej wykorzystywane serwisy w systemach klasy MES

Serwisy podzielone są na cztery główne nurty wymiany informacji, odpowiadające komunikacji z systemami wyższego poziomu, np. klasy ERP. Są to odpowiednio: definicja operacji, zdolności operacyjne, polecenia wykonywania operacji oraz odpowiedzi związane z realizacją operacji. Każdy z nurtów posiada własne serwisy, które wzajemnie na siebie wpływają lub które są realizowane we wzajemnym uzależnieniu. Są to:

- Zarządzanie definicjami, skupiające się formułowaniu zasad wytwarzania produktów w zależności od warunków lokalnych. Tworzone jest na etapie formułowania technologii produkcyjnej i reprezentuje projektowanie, utrzymywanie i aktualizowanie wiedzy, która musi zostać przekształcona w wymagania produkcyjne, dostarczane w postaci kart marszruty produktów, receptur czy technologii produkcyjnych.
- Zarządzanie zasobami, niezbędne do planowania produkcji w oparciu o szczegółowe wymagania dla poszczególnych produktów, jak również do

określania zdolności produkcyjnych w odniesieniu do zasad wytwarzania i w zależności od chwilowo dostępnych zasobów (materiałów, infrastruktury lub personelu).

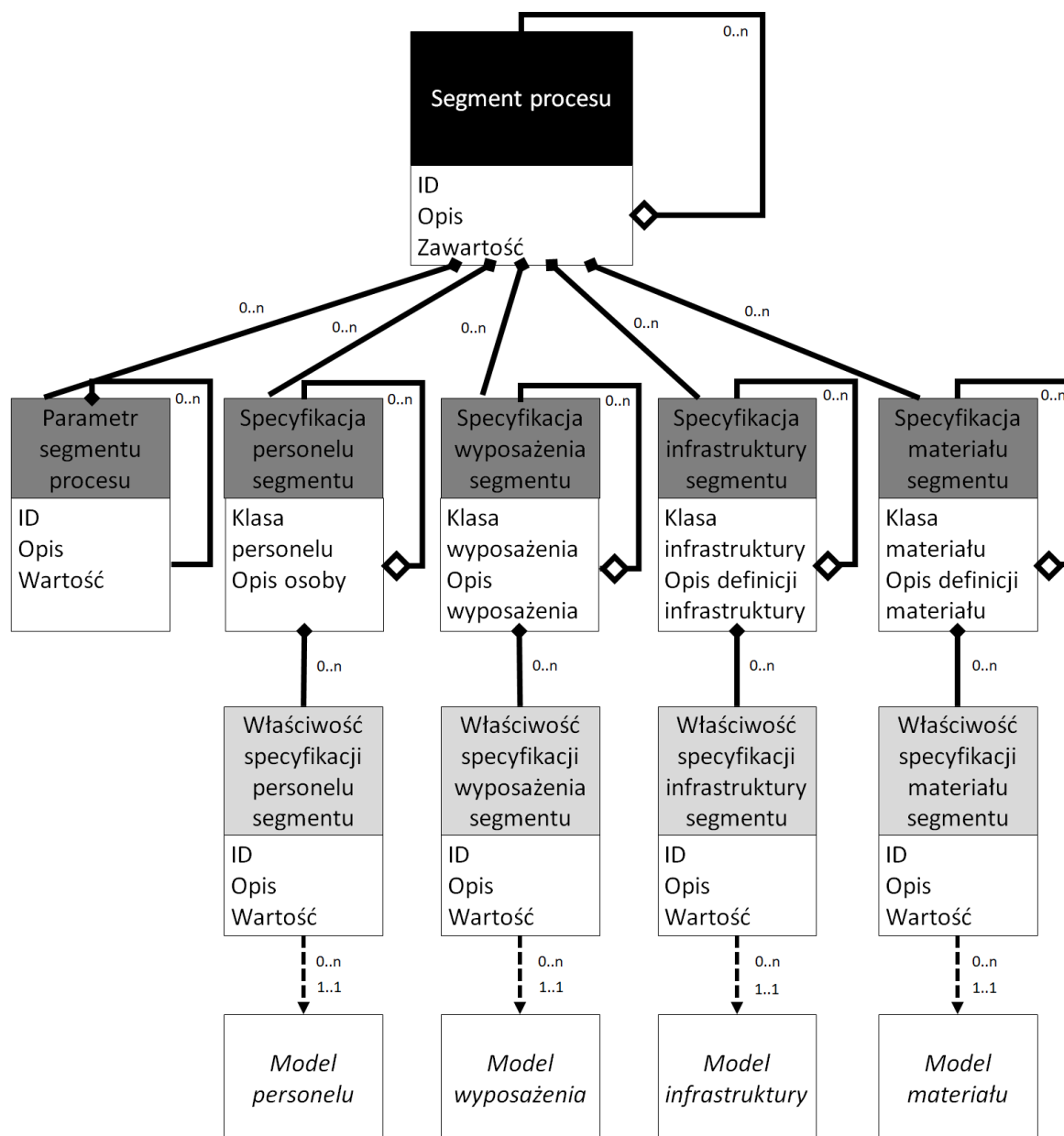
- Szeregowanie zadań produkcyjnych, wykorzystywane do optymalizacji zleceń z uwzględnieniem dostępnych chwilowo zasobów, zasad wytwarzania oraz produktów do wyprodukowania. Uwzględnia ponadto czynności niezwiązane bezpośrednio z produkcją, np. czasy przebrojeń, aktywności serwisowe i utrzymaniowe czy przerwy technologiczne wynikające z wykorzystania infrastruktury produkcyjnej.
- Asygnacja aktywności, związana z przekierowywaniem odpowiednich kroków lub zadań produkcyjnych do personelu technicznego, poszczególnych stanowisk wykonawczych lub innego sprzętu wykorzystywanego do produkcji. Asygnacja odbywa się w uzależnieniu od planu (uszeregowanych zadań produkcyjnych).
- Zarządzanie wykonywaniem produkcji, polegające na bezpośrednim zarządzaniu jednostkami wykonawczymi, np. personelem technicznym lub urządzeniami w celu zadawania poleceń produkcyjnych w postaci list pracy. Listy te, oprócz wymagań procesowych i nastaw, obejmują także materiały i partie produktów do wytworzenia. Efektem wykonywania produkcji jest produkt końcowy opisany definicjami.
- Gromadzenie danych, realizujące śledzenie procesów i gromadzenie danych produkcyjnych dla dwóch potrzeb: tworzenia historii produkcji i związania jej z wytwarzanym produktem oraz asygnacji zasobów i umożliwienia dynamicznego wpływania na produkcję, np. podczas wystąpienia błędów w wytwarzaniu i konieczności powtórzenia kroku produkcji lub przeprowadzenia naprawy produktu. Śledzenie danych obejmuje akwizycję sygnałów z systemów pomiarowych, danych procesowych z urządzeń wykonawczych i kontrolerów czy złożonych systemów wspomagania produkcji, np. systemów wizyjnych i identyfikacyjnych.
- Analiza efektywności, skupiająca się na badaniu wydajności produkcji i raportowaniu o niej. Prowadzona jest w oparciu o dane zebrane w kroku gromadzenia danych i wykorzystuje mechanizmy takie jak filtracja, agregacja czy przetwarzanie z wykorzystaniem różnych technik, np. statystyki czy

eksploracji danych. Efektem analiz są oceny efektywności produkcji, w tym określanie wpływu wykorzystywanych zasobów na proces produkcyjny.

- Śledzenie aktywności zasobów, mające na celu poprawę ich wykorzystania i usprawniające produkcję. Obejmuje monitorowanie stanu urządzeń i stanowisk produkcyjnych, obecność i lokalizację personelu technicznego oraz sposób i tempo wykorzystania materiałów. Śledzenie wspierać może analizę efektywności.

Model ANSI/ISA95 umożliwia opisanie tych serwisów, wraz ze szczególnymi składowymi systemów produkcji oraz samego procesu produkcyjnego. Modelowanie odbywać się może poprzez dekompozycję na mniejsze składowe, np. poprzez rozbicie procesu na mniejsze kroki, które mogą zostać zdefiniowane jako segmenty procesu (ang. Process Segment) i które mogą dotyczyć aktywności w obszarze ograniczonym lokalizacyjnie (np. w obszarze stacji produkcyjnej). Segment procesu cechować może się odpowiednimi właściwościami i składowymi w postaci personelu, wykorzystywanego sprzętu, infrastruktury i materiałów. Przykład segmentu procesu przedstawiony został na rysunku 2.2.3.

Zaprezentowany model segmentu procesu obejmuje fragment, w którym realizowana jest dana operacja produkcyjna. Modeluje on wszystkie czynniki niezbędne do przeprowadzenia fragmentu procesu, m.in. wymaganą infrastrukturę, urządzenia wraz z nastawami czy zużywane materiały. W skład segmentu wchodzi specyfikacje personelu (ang. Personnel Segment Specification), wykorzystywanych narzędzi produkcyjnych (ang. Equipment Segment Specification), infrastruktury i aktywów produkcyjnych (ang. Asset Segment Specification) oraz wymaganych zasobów i materiałów (ang. Material Segment Specification), wraz z odpowiadającymi im właściwościami (ang. Properties).



Rysunek 2.2.3. – przykład modelu segmentu procesu przygotowany w standardzie ANSI/ISA95

Każda ze specyfikacji posiada własne właściwości i parametry, np. urządzenie wykonawcze zawierać będzie odpowiednie nastawy do pracy, a materiał posiadać będzie ilość niezbędną do wykorzystania w danym kroku. Wykorzystane specyfikacje i ich właściwości wchodzi w skład szerszych modeli, np. specyfikacja konkretnego materiału stanowi część modelu materiału.

Choć modele definiowane przez ISA-95 są bardzo ogólne i mogą zostać zaadoptowane do w zasadzie każdego środowiska produkcyjnego, nie jest możliwe

zaimplementowanie ich w środowiskach wspierających produkcję masowo-dostosowaną. Wynika to z trzech głównych powodów:

- Standard ISA-95 zaprojektowany został jako jeden wspólny standard dla całej realizacji produkcji. W praktyce jednak, przez mnogość technologii, stosowanych systemów, różnorodność specyfik produkcyjnych i serwisów oraz przede wszystkim przez usługowo-zorientowaną komunikację pomiędzy nimi, podejście takie jest wysoko abstrakcyjne i nieaplikowalne, głównie przez trudność w dynamicznym stosowaniu modeli informacyjnych ISA-95 dla nowopowstających wariantów produkcyjnych.
- Implementacja przepływu informacji za pomocą specyfikacji B2MML stworzona została z myślą o produkcji wsadowej, a ze względu na jej charakter, nie nadaje się do wykorzystania w trybie online, który jest niejednokrotnie wymagany przez zmienność procesu i krótkoseryjność produkcji.
- Choć standard ISA-95 uwzględnia technologię produkcyjną, np. umożliwiając wprowadzenie odpowiednich atrybutów i parametrów wyposażenia czy infrastruktury produkcyjnej, to informacje te nie są w żaden sposób wykorzystywane później w procesie zarządzania produkcją, jej śledzenia czy utrzymania.

Współczesne systemy realizacji produkcji muszą się zmieniać tak, aby adaptować się do szybko i licznie wprowadzanych nowych wariantów produktowych, nadążać z kontrolą nad wytwarzaniem, umożliwiać śledzenie produkcji i wspierać zarządzanie jej utrzymaniem. Większość obecnie stosowanych systemów informatycznych w przemyśle nie jest w stanie spełnić tych wymagań, głównie przez brak zapewnienia odpowiedniej wymiany informacji pomiędzy składowymi i serwisami. Dodatkowo, coraz częstsze wiązanie wiedzy o technologii produkcji z cyklem życia produktu wymusza integrację danych pomiędzy obszarami o dotychczas słabo wykształconym przepływie informacji. Wreszcie, zmieniające się potrzeby rynkowe i rozwój technik informatycznych, w tym IoT, narzędzi wspomagania projektowania oraz technik utrzymania, zapewniają nowe możliwości wykorzystania danych i zasobów technologicznych z różnych etapów życia produktu. Konieczne jest więc rozwijanie nie tylko technik klasycznego wspomagania produkcji, np. szeregowania zadań produkcyjnych czy zarządzania wiedzą o produktach, ale również łączenia ze sobą danych procesowych z danymi technologicznymi tak, aby

obsługiwać elastyczne procesy wytwarzania produktów końcowych, w tym wytwarzanych w krótkich i wysoko indywidualizowanych seriach.

2.3. Komunikacja w przemysłowych systemach informatycznych wykorzystująca modele obiektowe

Komunikacja w przemyśle, rozumiana jako wymiana danych pomiędzy różnymi węzłami, zapewniająca operacyjność urządzeń i dostęp do różnych informacji, stosowana jest odkąd pojawiły się pierwsze urządzenia z interfejsami cyfrowymi. Funkcjonalnie, w zależności od stosowanych systemów kontrolujących procesy, wyróżnić można m.in. komunikację poziomą (np. pomiędzy sterownikami PLC) oraz komunikację pionową (np. pomiędzy systemem PLC i SCADA). W zależności od specyfiki stosowanych technologii, wykorzystywane są różne architektury sieci przemysłowych. Należą do nich:

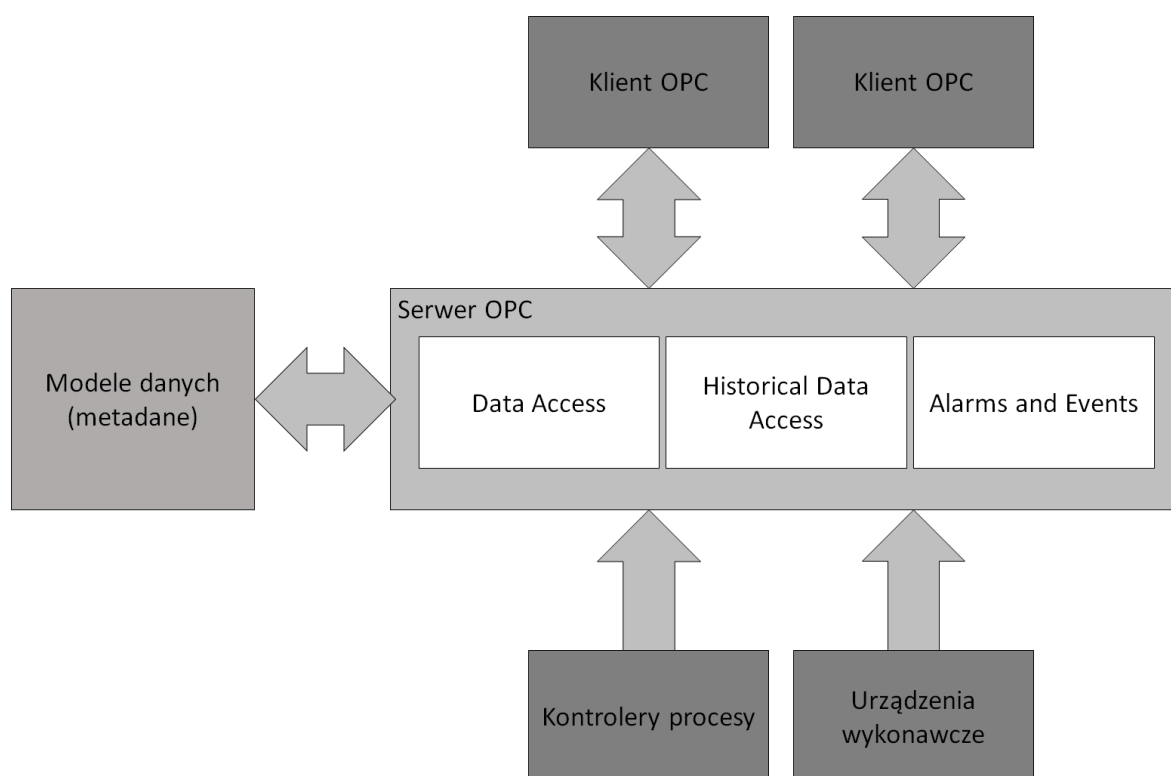
- Komunikacja punkt do punktu (ang. point to point), oparta o bezpośrednią komunikację pomiędzy dwoma węzłami. Przykładami tego typu architektury jest standard komunikacji szeregowej RS232 lub transmisja optyczna IrDA (ang. Infrared Data Association).
- Master-slave, w której komunikacja pomiędzy węzłami realizowana jest na zasadzie przydzielania dostępu do sieci przez jednostkę typu Master, która jednocześnie obsługuje medium transmisyjne pomiędzy jednostkami typu Slave. Przykładem tego typu architektury są sieci Profibus i Modbus.
- Klient-serwer, w której występuje podział ról pomiędzy serwerem, zapewniającym usługi dla łączących się klientów, zgłaszających różnego rodzaju żądania. Przykładem tego typu architektury jest serwer www, udostępniający różne zasoby w zależności od zapytań łączących się z nim klientów.
- Producent-Dystrybutor-Konsument, dla której wymiana danych odbywa się za pośrednictwem Dystrybutora, odpytującego w określonej kolejności Producentów o ściśle zdefiniowane dane i przekazującego te dane do odpowiednich Konsumentów. Przykładem sieci opartym o tą architekturę jest Factory Instrumentation Protocol (FIP).
- Token ring, w której transmisja danych przez węzły nie jest możliwa dopóki nie posiadają cyklicznie wymienianego w sieci żetonu (ang. token). Przykładem

wykorzystującym tą architekturę jest protokół 802.5 stworzony dla sieci LAN (ang. Local Area Network).

Wykorzystanie poszczególnych architektur w początkowych latach rozwoju technik komunikacyjnych w przemyśle polegało zwykle na tworzeniu przez producentów sprzętu i oprogramowania przemysłowego własnych implementacji, które w zależności od popularności komponentów przemysłowych zyskiwały większe lub mniejsze zastosowanie. Szybko jednak okazało się, że popularyzacja Internetu wykorzystującego w komunikacji podejście zorientowane usługowo spowodowała, że różnorakie dotychczas stosowane protokoły cechowały się niewielką skalowalnością i ograniczonymi możliwościami integracji [25]. Szczególnie widoczne stało się to podczas ery systemów SCADA, w których wystąpiła konieczność pionowej integracji kontrolerów procesu, np. sterowników PLC z systemami gromadzenia danych procesowych i rozwiązaniami HMI. Ponieważ mnogość stosowanych rozwiązań przemysłowych pokazała, że tworzenie dedykowanych interfejsów komunikacyjnych będzie niezwykle kłopotliwa w tworzeniu i utrzymaniu, konieczne było opracowanie nowego podejścia do wymiany danych w zmieniających się systemach przemysłowych. Odpowiedzią na to zapotrzebowanie było opracowanie standardu komunikacyjnego OPC DA (ang. OLE for Process Control Data Access) dla wymiany danych pomiędzy układami automatyki i systemami komputerowymi opartymi o system Windows, będącego wspólną inicjatywą konsorcjum składającego się z pięciu przedsiębiorstw: Rockwell Software, Opto 22, Intellution, Intuitive Technology oraz Fisher-Rosemount. Standard zapewnić miał możliwość ciągłej wymiany danych w sposób zorientowany usługowo, a dzięki swojej uniwersalności miał być dostosowany do wszystkich klas systemów stosowanych w przemyśle, od systemów kontrolnych po MES [26].

Ponieważ postępująca integracja pionowa w systemach przemysłowych umożliwiała opracowywanie kolejnych funkcjonalności wspierających produkcję (np. narzędzi bazodanowych, analitycznych czy zarządzania wiedzą), wymiana danych rozszerzona musiała zostać o informacje na temat sygnałów innych, niż tylko niezbędnych do bieżącej wizualizacji i kontroli procesu. Szczególnie dwa obszary wymagały rozwoju technik komunikacji: dostęp do informacji historycznych, archiwizowanych w postaci różnego rodzaju baz danych i plików z logami produkcyjnymi oraz wymiana danych do obsługi wydarzeń wspierających diagnostykę i utrzymanie instalacji produkcyjnych. Z tego

powodu popularny już wówczas interfejs OPC rozszerzony został o specyfikacje opisujące dostęp do danych historycznych – OPC HDA (ang. Historical Data Access) oraz definiujące interfejs do obsługi komunikatów diagnostycznych – OPC A&E (ang. Alarms and Events). Dzięki temu uzupełnieniu możliwe było nie tylko zapewnienie dostępu do danych procesowych pozyskiwanych w sposób ciągły i w trybie czasu rzeczywistego, ale również usługa dostępu do dużych zasobów zarchiwizowanych, niekoniecznie wymagających obsługi bieżącej oraz do danych pozyskiwanych w charakterze zdarzeniowym, na żądanie [27].



Rysunek 2.3.1. – schemat architektury OPC

Dalszy rozwój technologii, w tym upowszechnienie Internetu oraz coraz większe zapotrzebowanie na uniezależnienie komunikacji przemysłowej od platform sprzętowych spowodowało, że mało skalowalne rozwiązania klasycznego OPC stawały się niewystarczające do obsługi złożonych wymian danych pomiędzy systemami przemysłowymi. Konieczne było opracowanie podejścia możliwego do łatwej rozbudowy, integracji z różnymi typami systemów i operującego na różnych typach danych. Dodatkowo, w dobie rozwoju systemów zorientowanych usługowo, w których coraz istotniejszą rolę zaczęło odgrywać integrowanie danych procesowych z wiedzą o technologii, pojawiła się potrzeba powiązania pozyskiwanych danych

z odpowiadającymi im fragmentami technologii. Odpowiedzią na te zapotrzebowania był interfejs OPC UA (ang. Unified Architecture). Zapewnia on wymaganą skalowalność, wydajność, niezawodność i bezpieczeństwo danych, a jednocześnie umożliwia modelowanie danych w sposób zorientowany obiektowo. Modelowanie ma na celu uzupełnienie surowych danych procesowych o meta-informacje, dzięki którym możliwe jest wprowadzenie semantyki wskazującej na relacje pomiędzy danymi i pomiędzy generującymi je źródłami [28]. OPC UA stanowi warstwę pośredniczącą w komunikacji, która zdefiniowana jest na poziomie abstrakcji i która umożliwia zorientowane usługowo połączenie pomiędzy różnymi aktorami (urządzeniami, systemami, komponentami) [29]. Obecnie, interfejs ten ujęty jest wśród kluczowych rozwiązań technologicznych wspierających współczesną produkcję, a jego rozwój realizowany jest zarówno przez licznych partnerów przemysłowych jak również narodowe centra rozwojowe, np. US Industrial Internet Consortium czy niemiecką Plattform Industrie 4.0. OPC UA wykorzystywany jest jako narzędzie modelujące, które może być wykorzystane przy implementacji standardów i ogólnych modeli informacyjnych, np. PLCopen, OMAC czy ANSI/ISA-95. Podejście to poskutkowało opracowaniem modeli informacyjnych dedykowanych dla wybranych branż, np. AutomationML for OPC UA [30] czy OPC UA for ISA-95 Common Objects Model [31].

Architektura interfejsu OPC UA może być oparta o relację serwer-klient, w której serwer udostępnia wybrane informacje procesowe zgodnie z zapytaniami klientów, publisher-subscriber, w której serwer bezpołączeniowo transmituje dane do sieci, które mogą następnie być pobrane przez klientów oraz producent-konsumer, w której serwer pełni rolę dystrybutora danych. Same informacje zgromadzone są w oparciu o struktury opracowane uprzednio w postaci modelu informacyjnego. Dostęp do informacji oraz usług, które są związane z tymi danymi zapewniony jest przez zunifikowaną przestrzeń adresową (ang. Address Space) udostępnianą przez serwer. Korzeniem modelu jest węzeł główny OPC UA, będący abstrakcyjną klasą wskazującą na cechy poszczególnych źródeł informacyjnych. Każdy węzeł modelu identyfikowany jest przez unikalny identyfikator NodeID, stanowiący atrybut węzła. Dostęp do węzłów z zewnętrznych źródeł możliwy jest przez atrybuty BrowseName oraz DisplayName. Powiązania pomiędzy węzłami reprezentowane są przez referencje, które podobnie jak węzły zorientowane są obiektowo. Wszystkie węzły i referencje udostępnione są w przestrzeni adresowej OPC UA [26]. Przestrzeń ta może być przeglądana i wykorzystywana przez dowolnych klientów OPC

UA, łączących się do serwera. Typy referencji są znane dla poszczególnych klientów, wobec czego znane są również relacje wiążące poszczególne węzły. Każda referencja łączy dwa węzły, mogące być powiązane pośrednio lub bezpośrednio. Jednocześnie liczba referencji pomiędzy dwoma węzłami może być nielimitowana, możliwe jest więc modelowanie różnego rodzaju zależności. Klienci mogą wykorzystać usługi przeglądania przestrzeni adresowej w celu odkrycia pól z danymi poszczególnych węzłów oraz związanych z nimi informacji.

Interfejs OPC UA przewiduje osiem klas wykorzystywanych w modelowaniu:

- Obiekt (ang. Object) – klasa reprezentująca obiekty świata rzeczywistego, takie jak komponenty instalacji przemysłowej lub oprogramowania. Każda instancja tej klasy zdefiniowana jest przez odpowiedni typ i musi posiadać referencję do tego typu. Obiekty mogą być zagnieżdżone, tj. mogą tworzyć podgrupy oraz grupować zmienne, wywoływać metody i generować zdarzenia.
- Typ Obiektu (ang. ObjectType) – klasa wskazująca na typ obiektu, związana z obiektem przez referencję,
- Zmienna (ang. Variable) – klasa reprezentująca właściwości i wartości obiektów. Zmienne mogą zarówno reprezentować wartości statyczne, jak również zmieniać się w czasie. Wartości te mogą być odczytywane, zapisywane i subskrybowane przez klientów OPC. Muszą posiadać referencję do typu zmiennej oraz typu danej.
- Typ zmiennej (ang. VariableType) – klasa wskazująca na typ zmiennej, związana ze zmienną przez referencję,
- Typ danej (ang. DataType) – klasa wskazująca na typ wartości, związana z wartością zmiennej,
- Metoda (ang. Method) – klasa reprezentująca funkcjonalności realizowane lub wspierane przez węzły. Metody mogą posiadać właściwości opisujące argumenty wejścia i wyjścia wykorzystywane przez klientów OPC wywołujących dane metody.
- Powiadomienie o zdarzeniu (ang. EventNotification) – klasa wykorzystywana do określania informacji zwrotnej do klientów OPC o występującym zdarzeniu związanym z obiektem,

- Widok (ang. View) – klasa wykorzystywana jako punkt do reprezentacji zbioru węzłów przez klientów przeglądających przestrzeń adresową serwera OPC. Węzeł widoku określa punkt lub początek widoczności węzłów dostępnych w przestrzeni adresowej. Wykorzystywanie widoku wspierane jest przez mechanizmy filtrowania na podstawie typów referencji.

Komunikacja pomiędzy serwerem i klientami OPC realizowana jest za pomocą mechanizmu dostępu do danych (ang. Data Access). Każdy z klientów definiuje zbiór subskrypcji dla serwera, wykonywanych następnie w postaci sesji. Liczba monitorowanych elementów (węzłów) zależna jest od uzgodnień pomiędzy klientem i serwerem, związanych między innymi z ogólnymi limitami, aktualnym stanem zajętości łącza czy całkowitą liczbą wymaganych do zaktualizowania wartości. W ramach subskrypcji klient odpytuje o wymagane przez siebie dane zgodnie z narzuconym interwałem próbkowania, a serwer odsyła wymagane informacje jedynie w przypadku zajścia zmian w odpytywanych wartościach. Przesyłane dane uzupełniane są o informacje dotyczące statusu jakości i dwóch stempli czasowych: wygenerowania danych i udostępnienia przez serwer. Celem sesji jest grupowanie odpytywanych danych i optymalizacja ich wysyłania: odpowiedzi na zapytania nie są wysyłane natychmiastowo ale są najpierw grupowane i kolejgowane. Odpowiedzi serwera na zapytania klientów obejmują, oprócz transmitowanych danych, komunikaty notyfikacji (ang. Notification Messages) i wiadomości podtrzymania (ang. KeepAlive Messages) w celu poinformowania o aktywności serwera i przyjęciu zapytania do realizacji [32].

Oprócz mechanizmu dostępu do danych, OPC UA wspiera również mechanizmy archiwizowania i dostępu do danych zarchiwizowanych. Realizowane jest to za pomocą interfejsu dostępu historycznego (ang. Historical Access) związanego z konkretnym obiektem. Mechanizm ten umożliwia przechowywanie informacji o zmianach historycznych wartości danych atrybutów obiektu np. dla potrzeb przyszłego raportowania czy analityki. Dodatkowo wylicza wartości zagregowane, np. wartości minimalne i maksymalne, średnie czy mediany. Klienci OPC mogą odpytywać o dane historyczne za pomocą mechanizmu odczytu. Ponieważ OPC UA jest jedynie interfejsem komunikacyjnym, przechowywanie wartości historycznych musi być zrealizowane w zewnętrznych repozytoriach, np. relacyjnych bazach danych.

Ponieważ interfejs OPC UA powstał w celu zapewnienia uniwersalnej komunikacji pomiędzy różnymi rozwiązaniami stosowanymi w przemyśle, stanowi pewnego rodzaju standard możliwy do wykorzystania, zwłaszcza w architekturach wspierających nowoczesne wytwarzania, np. RAMI4.0. Poprzez właściwości przestrzeni adresowej, która w bardzo szczegółowy sposób modelować może fizyczną strukturę instalacji przemysłowej, stanowi on potencjalnie wartościowe narzędzie w wiązaniu danych procesowych z technologią produkcji, opisywanym przez drugą oś poziomą architektury RAMI4.0, co zostało wykorzystane w badaniach autora rozprawy i opisane w kolejnych rozdziałach.

2.4. Analiza skupień w eksploracji danych

Współczesne rozwiązania analityczne wykorzystują różnorakie techniki pozyskiwania informacji i budowania wiedzy. Jednym ze stosowanych do tego celu rozwiązań jest eksploracja danych, definiowana jako pozyskiwanie wzorców, powiązań lub prawidłowości w gromadzonych zbiorach w celu wydobywania ważnych lub ciekawych informacji [33]. Eksploracja danych jest dodatkowo procesem przetwarzania zbiorów danych z wykorzystaniem różnych perspektyw, tak, aby podsumować je w postaci nowo utworzonych informacji [34]. Ewolucja technik eksploracji danych nastąpiła wraz z rozwojem statystyki i badań nad sztuczną inteligencją, a do najczęściej stosowanych w niej podejść należą:

- klasyfikacje, pozwalające na kategoryzowanie elementów zbioru (obserwacji) w znane grupy lub klasy na podstawie atrybutów (cech) [35],
- klasteryzacje, służące do grupowania elementów zbioru w klastry bez znajomości docelowej liczby podziałów i oznaczeń (etykiet) tych grup [35],
- analiza asocjacji, służąca do wykrywania wzorców i powiązań pomiędzy danymi w analizowanych zbiorach [34],
- drzewa decyzyjne, umożliwiające klasyfikację obserwacji do znanych klas przez przyporządkowanie atrybutom odpowiednich decyzji [36],
- predykcje, które podobnie jak klasyfikacje umożliwiają przyporządkowanie obiektów do danych klas, jednak wykorzystują do tego celu mechanizmy przewidywania przyszłych wystąpień, np. trendów zmian i wartości atrybutów [37].

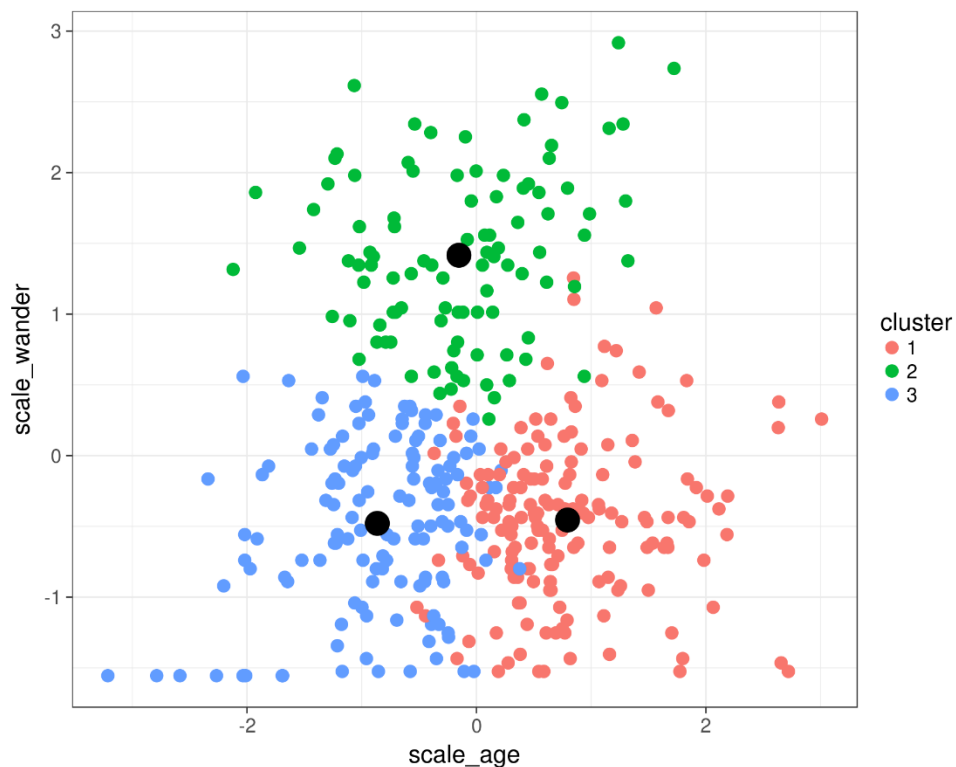
Eksploracja danych od lat znajduje zastosowanie w różnorodnych dziedzinach, m.in. rozpoznawaniu obrazów, mowy i pisma z wykorzystaniem systemów komputerowych, badaniach genetycznych, analizie rynków finansowych i operacji bankowych, rozpoznawaniu trendów w handlu czy prognozowaniu sprzedaży. Ze względu na upowszechnienie rozwiązań IoT i transformację systemów produkcyjnych, coraz częściej znajduje również zastosowanie w przemyśle do analizy danych produkcyjnych, jakościowych i związanych z funkcjonowaniem procesów produkcyjnych [38].

Jednym z podstawowych pojęć w eksploracji danych i uczeniu maszynowym jest analiza skupień, nazywana również klasteryzacją. Ma ona na celu podział zbioru obserwacji na grupy zgodnie z podobieństwem poszczególnych cech tych obserwacji i należy do grupy metod klasyfikacji nienadzorowanych (ang. *unsupervised*), w której przyporządkowanie poszczególnych obserwacji do grup realizowane jest bez wcześniejszego uczenia zbioru na próbkach treningowych. Dzięki podziałowi na grupy możliwe jest wyodrębnienie charakterystycznych cech wydzielonych podzbiorów, wyznaczanie różnic pomiędzy nimi, redukcja liczebności zbiorów danych i ich kategoryzacja lub wykrywanie ukrytych zależności wiążących obserwacje [39]. Do najpowszechniejszych dziedzin, w których stosowana jest analiza skupień należą: analiza obrazów, bioinformatyka, wykrywanie wzorców w różnego rodzaju danych, uczenie maszynowe oraz grafika komputerowa [40]. Pojęcie klasteryzacji wprowadzone zostało w 1939 roku przez Roberta Tryona w trakcie jego badań nad psychologią osobowości [41], a z czasem przeniesione zostało do dziedziny informatyki.

Ponieważ klasteryzacja to generalne zadanie podziału zbioru wejściowego na grupy, istnieją różne metody dokonujące tego podziału, uwzględniające stopień podobieństwa obserwacji, różnice pomiędzy poszczególnymi grupami lub gęstość dystrybucji cech w przestrzeni danych. Z tego powodu analiza skupień może zostać zdefiniowana jako optymalizacja wielokryterialna, w której poszukiwanie docelowego podziału zbioru na grupy zależne jest od wymagań problemu analitycznego, specyfiki grupowanych danych i właściwości zastosowanej metody podziału. Istnieje wiele różnych technik klasteryzacji, a ich podział dokonywany był przez autorów licznych źródeł literaturowych. Najbardziej ogólnym sposobem podziału metod klasteryzacji jest podział na techniki miękkie (ang. *soft clustering*) i twarde (ang. *hard clustering*), różniące się od siebie tym, że obserwacja może należeć do wielu (w przypadku miękkiego) i tylko do

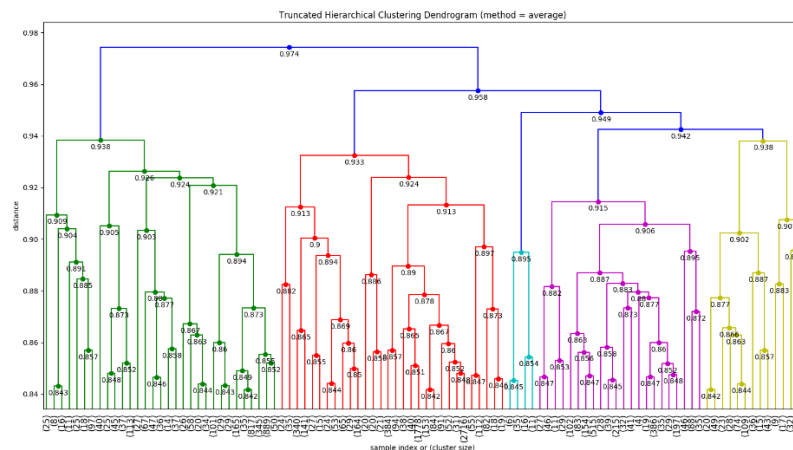
jednego klastra (w przypadku twardego) [42]. Bardziej szczegółowy podział technik klasteryzacji opisuje metody ze względu na połączenia pomiędzy obserwacjami (ang. connectivity-based), połączenia pomiędzy obserwacjami i środkami klastrów (ang. centroid-based), wykorzystujące modele rozkładu prawdopodobieństwa (ang. distribution-based) i hierarchiczne (ang. hierarchical) [43]. Obecnie, ze względu na mnogość dostępnych technik i specyfikę ich wykorzystania, najczęściej spotykanym podziałem metod klasteryzacji są podejścia opisane m.in. w [44] i [45]:

- Metody partycjonowania zbioru, bazujące na podziale na docelową, znaną liczbę grup na podstawie podobieństwa obserwacji wyrażanego za pomocą odpowiednich metryk np. odległości pomiędzy środkami grup i przyporządkowanych do nich obserwacji. Do zestawu tych metod należą m.in. algorytmy *k*-Means [46] czy PAM (ang. Partitioning Around Medoids) [47]. Podstawowymi zaletami tej grupy metod jest łatwość implementacji oraz niska złożoność obliczeniowa algorytmów, a przez to skalowalność do dużych zbiorów obserwacji. Niestety wymagają zwykle podania z góry docelowej liczby klastrów, a jakość podziału zbioru może silnie zależeć od warunków początkowych, w szczególności doboru punktów startowych wokół których budowane są klastry. Dodatkowo, nie posiadają zdolności do wykrywania wartości odstających, które w toku działania metod przyporządkowywane są do grup z pozostałymi obserwacjami.



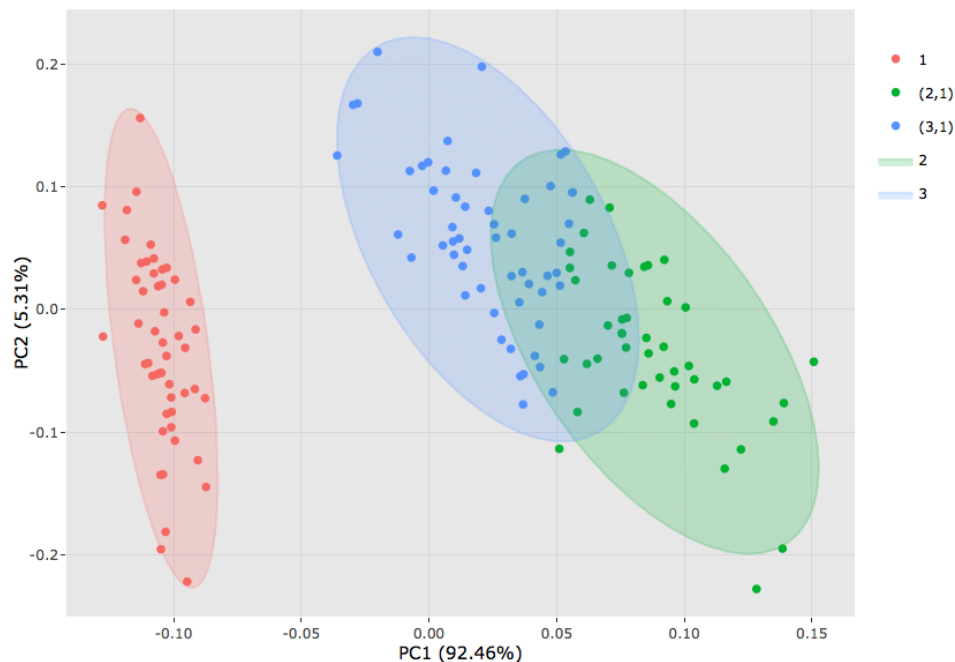
Rysunek 2.4.1. – przykład klasteryzacji metodą k -Means z zaznaczonymi klastrami i centroidami (źródło: rpubs.com)

- Metody hierarchiczne, które nie wymagają podawania z góry docelowej liczby grup, a które tworzą podziały zgodnie ze strukturą drzewa. Określanie przynależności obserwacji do grup realizowane jest przez wybór odpowiedniego poziomu drzewa (dendrogramu) na podstawie określenia stopnia podobieństwa obserwacji. Do najczęściej stosowanych metod tej kategorii należą podejścia aglomeracyjne (ang. bottom-up, od pojedynczych obserwacji) lub poprzez podziały (ang. top-down, od wszystkich obserwacji zbioru) [48]. Algorytmy z tej grupy nie wymagają podawania docelowej liczby grup i cechują się łatwością implementacji, niestety nie nadają się do operacji na dużych zbiorach danych ze względu na dużą złożoność obliczeniową. Jednocześnie są nieodporne na wartości odstające, które mogą być klasyfikowane jako osobne klastry.



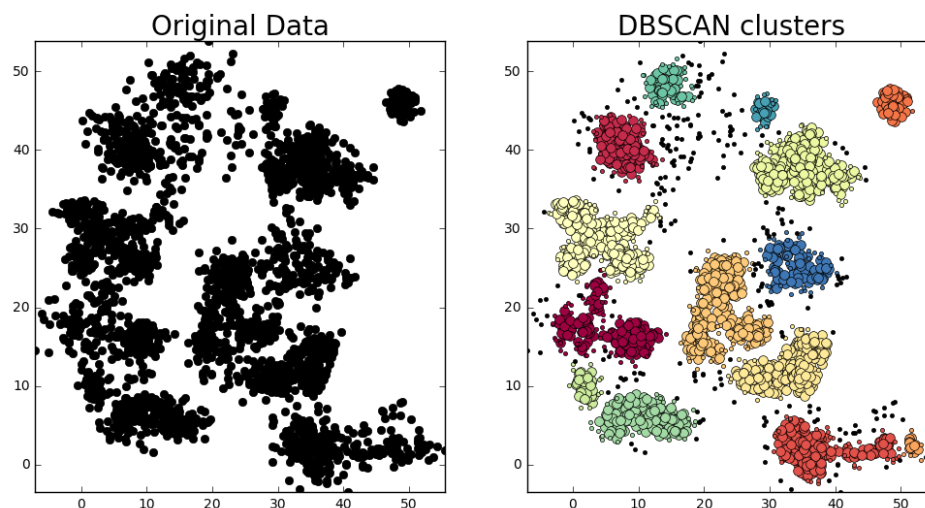
Rysunek 2.4.2. – przykład dendrogramu z połączeniami obserwacji
(źródło: stackoverflow.com)

- Metody rozmyte, umożliwiające przyporządkowanie obserwacji do więcej niż jednego klastra. Dopasowanie obserwacji do grupy definiowane jest z wykorzystaniem stopnia lub funkcji przynależności, typowo normalizowanych w przedziale od 0 do 1, choć niekoniecznie sumujących się do 1, co powoduje, że nie można ich traktować jako prawdopodobieństwa przydziału do danego zbioru. Do najczęściej stosowanych metod rozmytej kategoryzacji należą Fuzzy C-Means [49] oraz metoda FLAME (ang. Fuzzy clustering by Local Approximation of Member ships) [50]. Ponieważ jednak określanie tego stopnia wymaga dodatkowych obliczeń, algorytmy te są bardziej wymagające ze względu na złożoność obliczeniową.



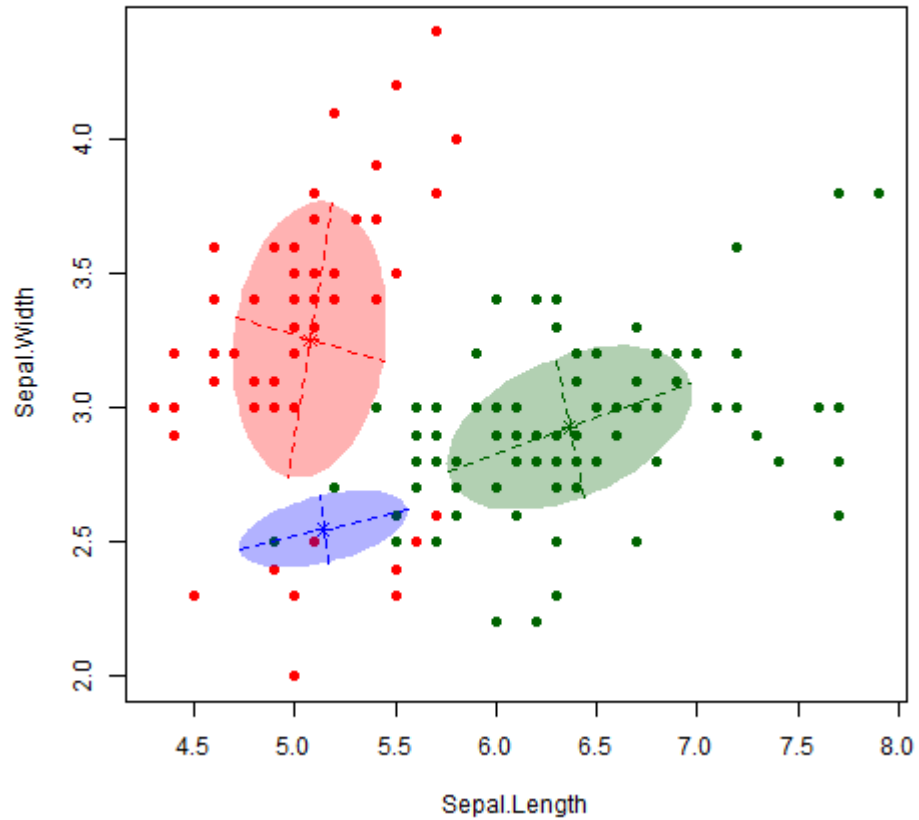
Rysunek 2.4.3. – przykład podziału zbioru z przynależnością obserwacji do dwóch klastrów (źródło: Autoplotly - Automatic Generation of Interactive Visualizations for Popular Statistical Results)

- Metody gęstościowe, w których zgrupowania tworzone są na podstawie zagęszczenia obserwacji w przestrzeni cech, a w których obserwacje wyraźnie odstające od grup traktowane są jako szum lub granice klastrów. Najczęściej wykorzystywanym algorytmem z tej grupy jest DBSCAN (ang. Density-Based Spatial Clustering of Applications with Noise) [51] oraz OPTICS (ang. Ordering Points to Identify the Clustering Structure) [52]. Algorytmy te nie wymagają podawania z góry liczby podziałów oraz cechują się zdolnością do wykrywania wartości odstających. Nie nadają się jednak do operacji na danych wielowymiarowych oraz formujących grupy o zmiennych stopniach gęstości obserwacji.



Rysunek 2.4.4. – przykład podziału zbioru z użyciem algorytmu DBSCAN
(źródło: Github.com)

- Metody bazujące na modelach, które stosowane do grupowania poprzez próbę odtworzenia wzorcowego modelu, z którego powstały obserwacje. Poszukiwane klastry oraz przynależność obserwacji do grup są w nich definiowane przez wyszukiwane modele. Przykładami tych metod są klasteryzacja wielowymiarowymi rozkładami normalnymi (ang. multivariate normal distribution [53] lub EM (ang. Expectation-Maximization) [54]. Podstawową zaletą tego typu algorytmów jest zdolność do przyporządkowania obserwacji do klastrów zgodnie z wyznaczonym prawdopodobieństwem, jednak przez konieczność budowy modelu są złożone obliczeniowo i wymagają dużej mocy przetwarzających.



Rysunek 2.4.5. – graficzna reprezentacja metody EM (źródło: lancaster.ac.uk)

2.4.1. Klasteryzacja metodą k -Means

Jednym z najpowszechniej stosowanych algorytmów partycjonowania zbioru, głównie ze względu na prostotę implementacji i skalowalność do dużych zbiorów danych jest algorytm k -Means, zaproponowany przez MacQueena w 1967 [55]. Algorytm zakłada iteracje na zbiorze danych z obserwacjami tak długo, aż osiągnięte zostanie kryterium stopu wyrażone np. w postaci maksymalnej dopuszczalnej liczby iteracji lub w postaci braku zmian w lokalizacji centroidów, definiowanych jako środki geometryczne wydzielanych grup. Do określenia dopasowania obserwacji do centroidów (środków) klastrów wykorzystywana jest najczęściej metryka euklidesowa, definiująca odległość obserwacji do centroidu, wyrażona jako:

$$d_E(x, r) = \sqrt{\sum_{j=i}^n (x_i - r_i)^2}$$

gdzie:

n = rozmiar wektora wejściowego (liczba cech obserwacji)

x_i = wartość i -tej cechy badanej obserwacji

r_i = wartość i -tej cechy centroidu

Oprócz metryki euklidesowej stosować można również inne metryki, np. odległość Manhattan:

$$d_M(x, r) = \sum_{j=i}^n |x_i - r_i|$$

lub kwadrat odległości euklidesowej:

$$d_{SE}(x, r) = \sum_{j=i}^n (x_i - r_i)^2$$

. W pierwszej iteracji algorytmu k -Means konieczne jest wskazanie docelowej liczby podziałów na k klastrów oraz wskazanie na centroidy tych klastrów (typowo ustalone losowo). Następnie przeprowadzany jest podział zbioru obserwacji na ustaloną liczbę k podziałów i obliczenie stopnia dopasowania obserwacji, np. za pomocą opisanej powyżej metryki euklidesowej. Po zgrupowaniu obserwacji konieczne jest wyliczenie środka ciężkości utworzonego klastra jako średniej arytmetycznej poszczególnych cech przypisanych obserwacji. Przemieszczenie centroidów w stosunku do pierwotnego zgrupowania oznacza konieczność ponownego dopasowania obserwacji do klastrów. Przypisanie obserwacji powtarzane jest więc tak, aby przyporządkować je do najbliższych centroidów. Kroki zmiany centroidów i przypisywania do nich obserwacji są powtarzane tak długo, aż osiągnięty zostanie warunek stopu, np. gdy centroidy przestaną się przemieszczać lub przemieszczać się będą na odległość mniejszą niż zdefiniowana parametrem. Ogólny opis implementacji algorytmu k -Means przedstawiony został w listingu 2.4.1.1.

Listing 2.4.1.1. – implementacja podstawowego algorytmu k -Means

1. Inicjalizacja k obserwacji początkowych stanowiących centroidy pierwszego podziału
2. Obliczenie odległości każdej z obserwacji do każdego centroidu zgodnie z zadaną metryką
3. Przypasowanie obserwacji do centroidów na podstawie ich minimalnej odległości
4. Wyliczenie nowych centroidów jako średniej arytmetycznej ze wszystkich obserwacji z danego klastra
5. Powtórzenie kroków 2., 3. i 4. tak długo, aż centroidy następujących po sobie podziałów zostaną przeniesione o odległość mniejszą niż zadana (osiągnięcie warunku stopu).

Efektem działania algorytmu jest podział zbioru na grupy z obserwacjami, dla których zminimalizowana zostaje metryka, np. opisana wcześniej suma kwadratów odległości obserwacji do centroidów swoich grup. Ze względu na swój charakter, metoda jest efektywna czasowo, posiadając złożoność obliczeniową dla każdej iteracji na poziomie $O(kN)$ (gdzie N to liczba obserwacji) oraz łatwa w implementacji. Jednocześnie wymaga podania a priori liczby podziałów k , jest wrażliwa na dobór centroidów startowych, jest nieskuteczna przy podziale zbioru zawierającego szumy i wartości odstające, może wpaść w minimum lokalne doprowadzając do nieprawidłowego podziału oraz nie posiada zdolności do identyfikacji zbiorów niewypukłych.

Dzięki popularności algorytmu k -Means, obecnie istnieje wiele jego wariantów i modyfikacji wprowadzonych w celu wyeliminowania negatywnych właściwości algorytmu, mogących wpływać na jakość grupowania. Problemy te były rozwiązywane w wielu dziedzinach i na wielu typach zbiorów danych, niekoniecznie tylko na danych rejestrowanych w środowiskach produkcyjnych. Rozwiązanie problemu wpadania w lokalne minima, prowadzące do nieprawidłowego podziału na grupy i problemu wyznaczania martwych centroidów, tj. takich, wokół których nie zostały zgrupowane żadne obserwacje, zaproponowane zostało w [56]. W publikacji tej, metoda nazwana przez autora jako Ruchome k -Means (ang. Moving k -Means) polega na wprowadzeniu dodatkowych mechanizmów oceny stopnia dopasowania obserwacji do centroidów w trakcie kolejnych podziałów opartych o metrykę euklidesową i dynamicznej zmianie dopasowania obserwacji do grupy w obrębie danej iteracji algorytmu. Zagadnienie

zniwelowania wpływu punktów początkowych rozpoczynających podział zbioru na klastry poruszone zostało i opracowane w [57]. Algorytm polega na przypadkowej selekcji k obserwacji ze zbioru początkowego, przypisaniu ich do k klastrów, a następnie stopniowemu dopasowywaniu kolejnych obserwacji i każdorazowym przeliczaniu centroidów klastrów w chwili zwiększania ich liczebności. W przypadku, gdy badana obserwacja znajduje się zbyt daleko od jakiegokolwiek klastra, dwa najbliższe ze sobą sąsiadujące klastry są łączone, a z oddalonej obserwacji tworzona jest nowa grupa. Innym wariantem metody jest k -Medoids, która zamiast metryki euklidesowej do oceny stopnia koncentracji klastra wykorzystuje faktyczne różnice pomiędzy wartościami poszczególnych cech obserwacji. Wynika to z faktu budowania środka geometrycznego klastra wokół rzeczywistej obserwacji (medoidu), a nie wyznaczanego za pomocą średnich arytmetycznych centroidu. Rozwiązanie to zaproponowane i opisane zostało w [58] i umożliwia lepszą interpretację fizyczną klastrów oraz ich środków, co przydaje się w specyficznych aplikacjach wymagających stosowania rzeczywistych wartości obserwacji. Jeszcze innym przykładem modyfikacji metody jest algorytm Fuzzy C-Means przedstawiony w [49], zakładający możliwość przypisania obserwacji do więcej niż jednego klastra na podstawie współczynnika dopasowania, przeliczanego po każdej iteracji, a którego zmiany na dostatecznie niskim poziomie wykorzystywane są do zatrzymania algorytmu. Istnieją dodatkowo liczne wariacje poszczególnych algorytmów, m.in. Z wykorzystaniem funkcji jądra umożliwiające przekształcenia cech do przestrzeni wymiarowych, opisane m.in. w [59] i [60] z wykorzystaniem prawdopodobieństwa do oceny obserwacji problematycznych, tj. odstających lub zbliżonych do wielu klastrów, przedstawione w [61] i [62] lub z zastosowaniem teorii zbiorów przybliżonych do oceny stopnia przynależności obserwacji do klastrów, opisane m.in. w [63] oraz [64].

Ceną prostoty rozwiązania i łatwości w implementacji algorytmu k -Means jest konieczność podania z góry liczby podziałów zbioru, a więc docelowej liczby klastrów, do których mają zostać dopasowane obserwacje. Istnieje kilka znanych podejść do automatyzacji wyznaczania liczby k . Jedną z najpopularniejszych metod jest metoda łokcia (ang. Elbow Method) polegająca na dzieleniu zbioru na różne liczby podziałów k od 1 do n i wyznaczaniu metryki wariancji poszczególnych klastrów wyrażonej np. jako stosunek wariancji pomiędzy grupami do całkowitej wariancji zbioru (średniej sumy odległości kwadratowych wewnątrz klastrów do sumy odległości kwadratowych całego zbioru) lub jako stosunek wariancji pomiędzy grupami do wariancji wewnątrz grup (F-test ANOVA).

Wyznaczone wariancje dla każdej liczby k wykreślone są na wykresie w funkcji k , a punkt przegięcia wykresu, nazywany punktem łokcia, w którym następuje wyłuszczenie wariancji na podstawie progowo ustalonej względnej różnicy wariancji pomiędzy kolejnymi k , definiuje optymalną liczbę podziału zbioru. Podejście opisane zostało m.in. w [65]. Innym sposobem automatycznego wyznaczania liczby k jest metoda profilu (ang. Silhouette Method), która oprócz wyznaczania średniej odległości wewnątrz grupy wyznacza także średnią odległość do najbliższego klastra. Dzięki temu możliwa jest ocena podobieństwa obserwacji do własnego klastra w porównaniu do centroidów innych klastrów. Typową miarą oceny sylwetki jest odległość euklidesowa lub odległość Minkowskiego, a optymalna liczba podziału zbioru ustalana jest na k , dla którego wartość wyniku sylwetki jest odpowiednio wysoka. Metoda opisana została m.in. w [66]. Kolejną możliwością jest zastosowanie kryterium informacyjnego, najczęściej Akaikiego (ang. Akaike Information Criterion, AIC) [67] lub Bayesowskiego (ang. Bayesian Information Criterion, BIC) [68], dokonującego oceny wyznaczonej liczby grup na podstawie porównania modeli wynikających z podziału na różne klastry. Algorytm działa dwuetapowo, najpierw dokonując podziału na odpowiednią liczbę klastrów kandydujących, a następnie oceniając wyznaczony podział z wykorzystaniem wskaźnika dopasowania do modelu, zależnego od wybranej metody (AIC, BIC lub innych). Stosować można również podejście oparte o walidację krzyżową, zaproponowane przez [69], w którym zbiór obserwacji dzielony jest na część poddawaną klasteryzacji z różnymi liczbami podziału na k klastrów, a drugi służy do walidacji podziału. Metoda zakłada stabilność wyznaczania grup, tj. tendencję do grupowania obserwacji w te same klastry przy prawidłowym dobraniu liczby k i jednoczesnym założeniu pracy na zbiorze z danymi, z których pierwotnie wyznaczona została liczba k . Metoda została rozwinięta na wiele sposobów, m.in. dla potrzeb zwiększania stabilności grup wynikającej z operacji na dużych zbiorach wejściowych, co opisane zostało w [70] i [71].

Technik automatycznego wyznaczania liczby k jest więcej, jednak zwykle cechują się one dużym stopniem złożoności, wynikającym albo z zastosowania dodatkowych mechanizmów oceny i selekcji, np. klasyfikacji lub operacji statystycznych, albo ze specyfiki operacji na dużych zbiorach danych w celu wypracowania wiarygodnych rezultatów. Z kolei najprostsza z powszechnych metod, metoda łokcia, nie zawsze wskazuje jasny punkt przegięcia lub może wskazać ich kilka, w efekcie będąc nieskuteczną w rzetelnym podziale zbioru na grupy. Z tego powodu dobór odpowiedniej techniki

automatyzacji wyznaczania liczby k silnie zależy od charakteru analizowanych danych, w szczególności ich ilości, wymiarowości i stopnia podobieństwa obserwacji. W przypadku operowania na danych produkcyjnych zbieranych podczas pracy w stałych, niezmiennych się warunkach produkcyjnych podstawowy algorytm k -Means będzie skuteczny. Jednak w dobie produkcji masowo-dostosowanej, w której dane mogą się dynamicznie zmieniać w zależności od różnych warunków pracy, produkowanych wariantów i występujących zjawisk zewnętrznych, właściwość stałej liczby podziałów algorytmu k -Means wyklucza go z analizy skupień takich danych i niezbędne jest automatyczne wyznaczanie tej liczby. Rozwiązania analityczne badające zagadnienia zmienności danych produkcyjnych nie mogą być budowane jako klasyczne systemy ekspertowe, zasilane wiedzą o technologii a priori, ale powinny samodzielnie rozpoznawać nowe grupy, np. typy wariantów produkcji czy anomalie oraz dostosowywać swoje mechanizmy do aktualnie panującego na produkcji stanu. Z punktu widzenia algorytmu grupowania cecha ta oznacza zdolność algorytmu do samodzielnego określania docelowej liczby klastrów i na to właśnie zagadnienie położony został nacisk w pracach badawczych autora niniejszej rozprawy.

2.5. Detekcja wariantów i profili produkcyjnych

Automatyczne wykrywanie wariantów i profili produkcyjnych w przemyśle jest zagadnieniem trudnym i skomplikowanym. Informacja o nastawach produkcyjnych typowo dostarczana jest do systemów automatyki w postaci odgórnej informacji z systemów kontroli produkcji, np. klasy MES, a układy wykonawcze realizują zadania produkcyjne jako odpowiedź na konkretne informacje pojawiające się na wejściu procesu. Zwykle jednak informacje z urządzeń wykonawczych i sygnały procesowe rejestrowane w trakcie pracy instalacji, np. zużywane energie lub tryby i statusy pracy, niekoniecznie dostarczają jednoznacznej odpowiedzi na temat realizowanego wariantu produkcyjnego. Wynika to głównie z charakterystyk wykonania instalacji, w których realizacja danych nastaw może znacząco różnić się ze względu na budowę stacji (np. długości przewodów doprowadzających sprężone powietrze) i stopień złożoności urządzeń wykonawczych (np. wzajemnego wpływu urządzeń, które mogą pracować jednocześnie).

Budowanie informacji o procesie, traktowanej jako odpowiedź układu automatyki na realizowany wariant produkcyjny, pozwala nie tylko dopasować indywidualny charakter instalacji do realizowanego profilu (jednakowo zbudowane stacje mogące produkować ten

sam typ produktu, jednak robić to w znacząco różniący się sposób, np. Z uwagi na lokalizację lub chwilowe warunki pracy), ale umożliwia dodatkowo dalsze wykorzystywanie tych informacji w celach diagnostycznych, np. dla potrzeb oceny jakości pracy stanowiska.

Istnieje kilka typów podejść do automatycznego wykrywania wariantów produkcyjnych w przemyśle, które zostały opisane w literaturze. Pierwszą, najczęściej spotykaną pulę rozwiązań stanowią podejścia przez analizę szeregów czasowych, umożliwiające wykrywanie właściwości pracy elementów infrastruktury poprzez badanie kształtów przebiegów sygnałów. Najczęściej stosowaną techniką jest tu klasyfikacja. Ponieważ zagadnienie analizy szeregów czasowych jest powszechnie wykorzystywane w różnych dziedzinach, nie tylko w przemyśle, w literaturze opisane zostały liczne rozwiązania wykorzystujące różne metody i algorytmy. Metody klasyfikacji szeregów czasowych pogrupowane zostały m.in. przez autorów publikacji [1]. Wyłonili oni następujące grupy metod klasyfikacji:

- Metody bazujące na porównaniu dwóch szeregów czasowych, referencyjnego i badanego,
- Metody bazujące na porównaniu wybranych interwałów z jednego lub wielu szeregów,
- Metody bazujące na wyszukiwaniu wzorca referencyjnego w całym przebiegu,
- Metody bazujące na analizie histogramowej częstości występowania powtórzeń fragmentów przebiegów,
- Metody kombinacyjne łączące jedną lub więcej powyższych technik,
- Metody bazujące na dopasowaniu modeli generacyjnych do szeregu czasowego, np. Z wykorzystaniem auto-regresji, ukrytego modelu Markova lub funkcji jądra.

Przebiegi czasowe mogą cechować się dużą złożonością i zmiennością, niejednokrotnie konieczne jest więc podzielenie na mniejsze fragmenty umożliwiające dalszą analizę. Istnieją opracowania sprawdzające możliwości segmentacji i uproszczenia szeregów czasowych, np. [72], w którym opisane zostały techniki analizy, takie jak ruchome okno, segmentacja metodą top-down lub desegmentacja metodą bottom-up i sposoby upraszczania przebiegów przez interpolację i regresję liniową. Inne techniki, np. zastosowanie aproksymacji z wykorzystaniem funkcji przedziałami liniowej przedstawione zostało w [73], a wykorzystanie algorytmów ewolucyjnych w [74].

Mnogość technik klasyfikacji przebiegów czasowych umożliwiła ich przebadanie oraz wzajemne porównywanie, np. autorzy pracy [75] porównali segmentację z wykorzystaniem perceptronów wielowarstwowych, konwolucyjnych sieci neuronowych i sieci rezydualnych. Podobne opracowania dla innych metod analizy przedstawione zostały w [76] oraz [77].

Część z ogólnych podejść wymienionych powyżej zastosowana została do wykrywania profili lub wariantów produkcyjnych w środowiskach przemysłowych. Przykładem budowania informacji o procesie jest tworzenie profili zużycia energii dla centrów obróbczych, w których wykorzystywane są metody wspomagające E-LI (ang. Electric-Load Intelligence), mające na celu poprawę zużycia energii. Przykład identyfikacji statusów pracy maszyny obróbczej i wykrywania stanów pracy jej urządzeń m.in. osi numerycznych, przedstawiony został w pracy [78]. Autorzy przeprowadzili dekompozycję informacji o zużyciu energii przez całą maszynę i szacowali zużycie energii przez poszczególne jej komponenty z wykorzystaniem metod klasyfikacji. Budowa klasyfikatora przygotowana została w etapie uczenia z wykorzystaniem maszyny wektorów wspierających (ang. Support Vector Machine, SVM), mających na celu rozpoznawanie i naukę sygnatur energetycznych poszczególnych komponentów. Informacja ta posłużyła następnie w etapie rozpoznawania aktywacji tych komponentów w trakcie bieżącej pracy maszyny. Zastosowanie wnioskowania warunkowego opartego o algorytm średniego odchylenia bezwzględnego (MAD) umożliwiło identyfikację statusów pracy stanowiska i zbudowanie informacji o zużyciu dla różnych wariantów obrabianych produktów (trybów pracy wrzecion). Innym przykładem jest podejście zaproponowane w [79], dedykowane dla centrów obróbczych, polegające na analizie czasowej zużycia energii, selekcji cech do analizy z wykorzystaniem statystycznej metody analizy głównych składowych (ang. Principal Component Analysis, PCA) i budowaniu klasyfikatora o obciążeniu osi z wykorzystaniem algorytmu K najbliższych sąsiadów (ang. K -Nearest Neighbours, KNN). Zidentyfikowane i skategoryzowane tryby pracy maszyny wykorzystywane były w etapie analizy sygnałów napływających w trakcie bieżącej pracy stanowiska i posłużyły do wyznaczenia informacji o zużyciu energii podczas obróbki różnych wariantów produktów. Jednym z problemów nakreślonych podczas analizy przebiegów czasowych jest nadmierne skumulowanie wpływu różnych stanów pracy maszyny w postaci przebiegów energii, które nie są możliwe do dekompozycji, a w efekcie do skutecznego wykorzystania. W pracy [80], autorzy skupili się na opracowaniu metody analizy przebiegu zużycia energii

przez komponenty przemysłowe i zaproponowali podejście do wydobywania informacji o chwilowych trybach pracy urządzeń z wykorzystaniem klasteryzacji fragmentów przebiegu, w których występują wartości szczytowe chwilowego zapotrzebowania na moc urządzeń. Samo podejście służy do określenia stanów pracy obrabiarki CNC, występujących podczas pracy urządzenia. Drugim z zaobserwowanych problemów jest złożoność danych na wejściu algorytmów prowadzących ocenę i wykrywanie wariantów lub profili produkcyjnych. Dostrzeżony on został przez autorów opracowania [81], którzy zaproponowali metodę analizy zagregowanych danych wejściowych w celu uproszczenia dalszych kroków analitycznych. Przebieg wejściowy zużycia energii przez obrabiarkę numeryczną przekształcany jest w uproszczony przebieg z wykorzystaniem estymatora jądrowego gęstości (KDE). Następnie wybrane fragmenty przebiegu identyfikowane są przez specjalistów zaznajomionych z technologią wytwarzania, a wartości charakterystyczne opisujące dane przebiegi i odpowiadające stanom pracy maszyny (np. stromości zboczy, maksima i minima lokalne, długości trwania odpowiedniego zużycia) wprowadzane są do algorytmu statystycznego w postaci nastaw. System następnie dokonuje automatycznej identyfikacji stanów pracy maszyny na podstawie przebiegów bieżących i dokonuje oceny jakości pracy poprzez dopasowanie do parametrów referencyjnych opisujących poszczególne stany. Podejściem przez wykorzystanie klasyfikatora lasu losowego oraz zwiększania gradientu jest rozwiązanie opisane w [82]. Dane o produkcji poszczególnych wariantów mikroelektroniki PCB gromadzone są na różnych etapach produkcyjnych w postaci nastaw i rezultatów z inspekcji (np. systemów wizyjnych), a następnie służą do nauki klasyfikatorów, wykorzystywanych w produkcji bieżącej do rozpoznawania wariantów i wykrywania nieprawidłowości w prowadzonym procesie produkcyjnym. Podejście to zakłada uproszczenie danych na wejściu klasyfikacji poprzez redukcję ich liczby tylko do najważniejszych składowych, wyznaczanych empirycznie lub określanych przy wsparciu technologów zaznajomionych ze specyfiką pracy instalacji. Przykładem dwuetapowego przetwarzania szeregów czasowych, w którym najpierw dane są odpowiednio filtrowane, a następnie przetwarzane jest rozwiązanie przedstawione w [83]. Surowe dane wejściowe o energii zużywanej w procesie formowania wtryskowego filtrowane są w celu usunięcia szumów pomiarowych. W tym celu, w oparciu o znajomość technologii, zbudowany został referencyjny model procesu wtryskiwania z wykorzystaniem automatu skończonego (FSM) oraz zastosowany filtr Savitzkiego-Golaya do wygładzenia segmentów przebiegu wydzielonych na podstawie

FSM. Odfiltrowane i zredukowane dane wejściowe analizowane są następnie przez sieć neuronową, mającą na celu identyfikację stanów aktywności maszyny.

Oprócz klasyfikacji szeregów czasowych do ich analizy stosowane są techniki klasteryzacji. Autorzy publikacji [84] zaprezentowali podejście do analizy przebiegów czasowych poprzez segmentację profilu zużycia energii zasilającej stanowisko obróbcze. Przebieg poddany został analizie PCA, mającej na celu wybranie próbek pomiarowych do dalszych analiz. Następnie wydzielone zostały segmenty odpowiadające charakterystycznym stanom definiującym pracę instalacji. Segmenty przeanalizowane zostały ze względu na wypłaszczenia, a następnie te odpowiadające stanom aktywności maszyny poddane zostały klasteryzacji metodą HDBScan tak, aby zgrupować odpowiednie profile w odpowiadające im stany pracy stanowiska. Rozwiązaniem prezentującym podejście do rozpoznawania typów wytwarzanych produktów na przykładzie maszyny CNC jest algorytm opracowany i opisany przez autorów w pracy [85]. Informacja o zużyciu energii wyrażona w postaci przebiegu czasowego poddawana jest klasteryzacji metodą *k*-means. Następnie wyłonione klastry analizowane są w celu wyłonienia okresów aktywności stanowiska, a dalej zasilone informacją o całkowitym czasie pracy spędzonym przy obróbce konkretnego produktu w celu wyłonienia okresu, w którym wykryta energia oznacza zapotrzebowanie na produkcję. Autorzy w tej samej pracy przedstawili porównawczą metodę wykrywania profili z wykorzystaniem analizy kroków podstawowych (BSA), służących do wyznaczenia zapotrzebowania na energię w poszczególnych krokach procesu. Kroki te są następnie identyfikowane na podstawie sygnałów z bieżącej pracy instalacji, a informacje o produkowanych wariantach budowane są przez odwzorowanie profili energetycznych odpowiadających odpowiednim krokom.

Drugim podejściem stosowanym do wykrywania profili pracy instalacji jest analiza spektralna. W dziedzinie przetwarzania sygnałów procesowych stosowana jest do określania składowych mocy sygnału przynależącego do danego przedziału częstotliwości. W przemyśle zagadnienia związane z analizą widmową typowo wykorzystywane są do badania powierzchni badanych przedmiotów, profilometrii i profilografii, a przez to pośrednio do detekcji i oceny produkowanych wariantów. Często metody te wykorzystywane są do jednoczesnego badania defektów powierzchni i służą do wsparcia mechanizmów diagnostycznych i kontroli jakości. Budowanie informacji o profilach okrągłości obrabianych przedmiotów w dziedzinie szlifowania przedstawione zostało w pracy [86]. Autorzy opracowali metodę analizy fragmentów przebiegów sygnałów

akustycznych w domenie częstotliwości, która posłużyła do oceny profilu okrągłości szlifowanego elementu. Podobna technika zastosowana została w podejściu do badania chropowatości obrabianego elementu, zaprezentowana w [87]. W opracowaniu [88] autorzy przeprowadzili klasyfikację błędów występujących podczas pracy pomp wirowych. Dane w postaci wibracji i natężeń prądów zgromadzone w trakcie różnego rodzaju zarejestrowanych błędów w pracy urządzenia wykonawczego przekształcone zostały do dziedziny częstotliwości i poddane klasyfikacji z użyciem maszyny wektorów wspierających. Analogiczne podejście, które oprócz maszyny wektorów wspierających wykorzystuje również sztuczne sieci neuronowe przedstawione zostało w [89]. Przykładem zastosowania analizy emisji akustycznej do oceny profili pracy drukarki 3D jest metoda zaprezentowana w [90]. W pierwszym etapie autorzy zbudowali model zdolny do określania stopnia porowatości drukowanych obiektów z wykorzystaniem konwolucyjnych sieci neuronowych, a następnie wykorzystali go do oceny bieżącej pracy urządzenia produkcyjnego. O ile jednak techniki analizy w domenie częstotliwości z powodzeniem stosowane są do wykrywania właściwości cechujących indywidualne produkty lub pracę specyficznych urządzeń wykonawczych, ze względu na specyfikę pomiarową skupioną na jednym punkcie instalacji, nie są stosowane do oceny wariantów produkcyjnych złożonych instalacji przemysłowych, np. całych stanowisk wyposażonych w wiele urządzeń.

Trzecim, rzadziej wykorzystywanym typem podejścia do wykrywania profili produkcyjnych i wariantów produkcji jest stosowanie technik statystycznych. Przykładem jest rozwiązanie opisane w [91], w którym autorzy przeprowadzili empiryczną ewaluację wartości średniego zapotrzebowania na energię w różnych trybach pracy maszyny, a następnie wykorzystali mechanizmy progowania do oceny bieżącego przebiegu czasowego tak, aby przez umożliwić dopasowanie segmentów przebiegu do zużycia i ocenić stany pracy stanowiska. Podejście to jest zbliżone do rozwiązania zaprezentowanego w [92], w którym autorzy budują informację o profilach produkcji w oparciu o dane czasowe zużycia energii. Energie wyznaczone są empirycznie dla pojedynczych kroków procesu montażu karoserii samochodowych, a następnie wykorzystywane są do budowania profilu zużycia energii związanego z produkcją konkretnego wariantu produktu końcowego. Profile związane są nie tylko z pracą przy konkretnym wariantcie, ale również z realizacją na konkretnym stanowisku montażowym, zgodnie z zaproponowaną przez autorów filozofią Methods-Energy Measurement,

zakładającą powiązanie referencyjnych informacji o pracy stanowiska z technologią jego wykonania.

Wszystkie zaprezentowane rozwiązania z powodzeniem zostały uruchomione i zwalidowane w warunkach laboratoryjnych lub przemysłowych, jednak ich podstawowymi ograniczeniami są: duży stopień skomplikowania danych na wejściach algorytmów (wyrażonych jako przebiegi czasowe lub częstotliwościowe sygnałów procesowych, zwykle chwilowej energii elektrycznej lub mocy), niejednokrotną konieczność dodania dedykowanych układów pomiarowych umożliwiających zebranie odpowiednich danych do analiz (np. układów emisji i detekcji dźwięku), budowanie informacji w oparciu o lokalny pomiar wielkości fizycznych (np. analiza wybranego fragmentu wytwarzanego produktu) oraz fakt zastosowania dla przypadków realizujących tylko jeden krok produkcyjny lub będących tylko w jednym trybie pracy jednocześnie. Dodatkowo metody opisane w literaturze nie skupiają się na wykrywaniu wariantów produkcyjnych samych w sobie, a raczej na wykrywaniu profili energetycznych lub wybranych właściwościach produktów końcowych, które odpowiadać mogą stanom pracy maszyny realizującej procesy wytwórcze pod różnymi warunkami (np. podczas pracy przy różnych wariantach produktów lub podczas różnych typów błędów), zwykle z ograniczeniem do identyfikacji ogólnych trybów pracy (np. aktywności, wyłączenia, oczekiwania na produkcję) bez rozróżnienia szczegółowych faz produkcji. Co więcej, tylko w jednym opracowaniu [92] podjęty został problem potencjalnych różnic w realizacji produkcji tego samego produktu na dwóch różnych stanowiskach. We wszystkich innych przypadkach badania prowadzone były dla pojedynczej (a więc lokalnej) instalacji, bez analizy wpływu czynników zewnętrznych na różne instalacje wykonujące te same czynności. W celu uzyskania bardziej precyzyjnej informacji związanej z realizacją różnych wariantów produkcyjnych przez złożone i różniące się od siebie stanowiska konieczne jest więc przeprowadzenie dalszych badań, a zakres prac przedstawionych w niniejszej rozprawie jest częściową odpowiedzią na to zapotrzebowanie.

2.6. Klasteryzacja danych zbieranych w przemysłowych systemach informatyki

Współczesne systemy informatyczne wykorzystywane w przemyśle wymagają gromadzenia i przetwarzania ogromnych ilości danych. Choć pierwsze techniki, które służyły do tych celów stosowane były w czasach, gdy celem tego gromadzenia

i przetwarzania była archiwizacja i dokumentacja procesów, to obecnie dane poddawane są różnorodnym operacjom mającym na celu poszukiwanie pól do optymalizacji efektywności, poprawę wydajności, analizę jakości czy wczesne wykrywanie anomalii. Jedną z ważniejszych kwestii w realizacji tych zadań eksploracji danych jest dobór odpowiednich technik analitycznych. Jednocześnie samo przygotowanie danych do wykorzystania przez te techniki bywa skomplikowanym i czasochłonnym zagadnieniem. W zależności od typu problemu do rozwiązania i specyfiki analizowanych danych, stosować można różne metody z zakresu klasyfikacji, grupowania, poszukiwania wzorców czy zależności. Chociaż istnieje wiele różnych metod grupowania, jedną z najpopularniejszych i najłatwiejszych do wdrożenia jest metoda *k*-Means, opisana w rozdziale 2.2, a która znalazła zastosowanie w wielu rozwiązaniach przemysłowych.

W pracy [93] problem planowania obłożenia stanowisk produkcyjnych uzupełniony został o analizę grup produktowych z wykorzystaniem metody *k*-Means. Produkty zgrupowane zostały ze względu na podobieństwa w technologii wytwarzania, co umożliwiło optymalizację szeregowania zadań i ustalania planów produkcyjnych. Zagadnienie związane z redukcją liczby wariantów produkcyjnych wytwarzanych w przemyśle metalowym opisane zostało w [94]. W przedstawionym rozwiązaniu dzięki klasteryzacji możliwe było wyłonienie najważniejszych grup produktowych i próba optymalizacji planowania zleceń produkcyjnych, które ze względu na zbyt dużą liczbę i złożoność nie mogły zostać bezpośrednio wyznaczone z wykorzystaniem modeli dyskretnych (ang. Discrete-Event Simulation, DES). Zasilenie modelu zredukowaną liczbą wariantów umożliwiło modelowanie m. in. planowania zapotrzebowania materiałowego. Liczne opracowania opisują wykorzystanie algorytmu *k*-Means do wykrywania nieprawidłowości w pracy urządzeń produkcyjnych. W pracy [95] przedstawiona została metodologia wyznaczania grup komponentów potrzebnych do harmonogramowania zadań produkcyjnych zgodnych z filozofią rekonfigurowanych systemów produkcji. Rozwiązanie problemu opracowano z wykorzystaniem zmodyfikowanej aglomeracyjnej hierarchicznej metody *k*-Means, dokonującej podziału na małe grupy rozpoczynając od całego zbioru i sukcesywnie dodając kolejne obserwacje do nominalnie utworzonych klastrów na podstawie warunku maksymalnej odległości euklidesowej. Otrzymane wyniki umożliwiły zgrupowanie komponentów oceniane na podstawie ich stopnia podobieństwa, a następnie ustalenie planów produkcji tak, aby doprowadzić do jak najmniejszej liczby przebrożeń stanowiska. Przykładem wykorzystania algorytmu *k*-Means do selekcji typów

materiałów używanych w produkcji ze względu na ich najbardziej optymalne zużycie w zależności od bieżących warunków panujących w środowisku pracy jest rozwiązanie przedstawione w [5]. Rezultaty z działania tego algorytmu zestawione zostały i porównane z efektami innych technik klasyfikacji, m.in. sieci neuronowych i drzew decyzyjnych.

Oprócz zagadnień analizowania i planowania produkcji, klasteryzacja stosowana jest do grupowania danych procesowych gromadzonych w trakcie produkcji. W publikacji [6] zaprezentowano metodę analizy danych historycznych z produkcji urządzeń mikroelektroniki opartą o klasteryzację metodą k -Means, test Kruskala-Wallisa i algorytm drzew decyzyjnych. Metoda posłużyła do analizy błędów funkcjonalnych mikroelektroniki wykrytych na etapie testów końcowych na liniach produkcyjnych, a które mogły wynikać z zastosowania różnych kombinacji stosowanych komponentów lub z różnego sposobu prowadzenia procesu technologicznego, np. Z wykorzystaniem konkretnych urządzeń produkcyjnych. Autorzy pracy [96] zaprezentowali metodę wykrywania wad i zużycia łożysk tocznych z wykorzystaniem modelu wibracyjnego, który poprzez analizę spektralną drgań dostarcza informacji o charakterystyce pracy składowych łożyska, poddawanych następnie dwuetapowemu procesowi klasteryzacji k -Means w celu określenia wad (uszkodzenia zewnętrznego lub wewnętrznego pierścienia). Podobnym przykładem analiz jest detekcja błędów w pracy wentylatora wyciągowego, prowadzona z wykorzystaniem różnych technik klasteryzacji, w tym k -Means i opisana w [97]. Inny przykład detekcji błędów produkcyjnych zaprezentowany został w opracowaniu [98], w którym autorzy analizowali sygnały dźwiękowe generowane podczas pracy drukarek 3D, tworzyli z nich przebiegi w domenie czasu i częstotliwości, a następnie poddawali metodzie klasteryzacji k -Means w celu wykrywania nieprawidłowości w pracy ekstrudera. Analiza błędów jakościowych odlewów aluminiowych przeprowadzona została z wykorzystaniem metody k -Means i opisana w pracy [99], w której autorzy analizowali wpływ danych procesowych gromadzonych podczas odlewania na jakość odlewu końcowego. W wielu przypadkach dostosowanie algorytmu do rodzaju gromadzonych danych doprowadziło do uzyskania dokładniejszych wyników w porównaniu z zastosowaniem algorytmu w jego podstawowej wersji. Bardziej szczegółowy opis podejść do wykrywania anomalii w przemyśle z wykorzystaniem m.in. klasteryzacji metodą k -Means przedstawiony został w rozdziale 2.7.

Klasteryzacja może być wykorzystywana w analizach rynkowych i opracowaniach statystycznych związanych z działaniem poszczególnych branż. Przykładowo,

w opracowaniu [100] zaprezentowane zostało wykorzystanie klasteryzacji metodą *k*-Means do zgrupowania przedsiębiorstw tureckich działających w branży obróbki drewna i powiązania ich w odpowiednie regiony gospodarcze na podstawie lokalizacji. Opracowanie [101] przedstawia wykorzystanie algorytmu *k*-Means do grupowania produktów sprzedażowych zgodnie z ich chwilowym zapotrzebowaniem rynkowym. Celem badań było wsparcie zarządzania stanami magazynowymi i zapewnienie ciągłości dostaw do punktów dystrybucyjnych. W pracy [102] przeprowadzono analizę danych sprzedażowych w celu wyłonienia grup produktowych cechujących się różnymi poziomami zainteresowań kupujących tak, aby zaplanować odpowiednie akcje produkcyjno-sprzedażowe. Z kolei w pracy [103] wykorzystano algorytm *k*-Means do analizy i podziału poszczególnych branż przemysłowych ze względu na zapotrzebowanie i absorpcyjność personelu produkcyjnego na przestrzeni różnych lat, określając w ten sposób rynki rozwijające się bardziej i mniej dynamicznie.

Zaprezentowane podejścia bazują przede wszystkim na podstawowej metodzie *k*-Means, w której docelowa liczba podziałów definiowana i podawana jest odgórnie. Niestety, ponieważ produkcja masowo-dostosowana cechować się może krótkimi seriami produkcji dla produktów oraz ich dużą wariantowością, dane pozyskiwane w jej trakcie tworzą niewielkie zbiory o różnym charakterze. Konieczne jest więc opracowanie podejścia, które nie tylko będzie zdolne do automatycznego wykrywania liczby wariantów produkcyjnych na małych zbiorach danych, a więc zdolnego do automatyzacji wyznaczania liczby podziałów zbioru na *k* grup, ale również radzącego sobie z wykrywaniem i odfiltrowywaniem okazjonalnych błędów produkcyjnych, charakterystycznych dla produkcji często zmieniającej się, a które reprezentowane są przez obserwacje odstające. Potrzeby te skłoniły autora rozprawy do przeprowadzenia badań nad automatycznym wyznaczaniem wariantów produkcyjnych z wykorzystaniem algorytmu *k*-Means i modyfikacji go tak, aby w prosty sposób wyznaczać automatycznie liczbę podziałów.

2.7. Wykrywanie anomalii produkcyjnych z wykorzystaniem technik eksploracji danych

Zagadnienie wykrywania błędów produkcyjnych jest jednym z podstawowych elementów poprawiania jakości produkowanych wyrobów i zwiększania efektywności produkcji. Mniej awaryjne stanowiska oznaczają produkcję bardziej płynną, szybszą i wydajniejszą, zmniejszając dodatkowo występowanie błędów funkcjonalnych mogących wpływać na prawidłowość działania produktów końcowych oraz ograniczając generowanie odpadów materiałowych. Stąd też niezbędne jest utrzymywanie ciągów technologicznych w jak najlepszej kondycji, a jednym z tego elementów jest budowanie wiedzy o anomaliami, jakie mogą występować w trakcie pracy instalacji. Wykrywanie anomalii odbywać się może na dwa sposoby:

- Poprzez opracowanie systemu uwzględnionego już na etapie projektowania stanowiska produkcyjnego lub poprzez doposażenie stanowiska w odpowiednie urządzenia zdolne do bieżącego pomiaru chwilowych wielkości fizycznych i parametrów procesu, ich analizy i wykrywania nieprawidłowości,
- Poprzez wykorzystanie istniejącej infrastruktury produkcyjnej bez jej modyfikowania, bazującej na wykorzystywaniu dostępnych wielkości fizycznych i parametrów procesu, ich przetworzeniu, zagregowaniu i ewentualnym uzupełnieniu tak, aby budować cyfrowy obraz procesu i wnioskować na jego temat.

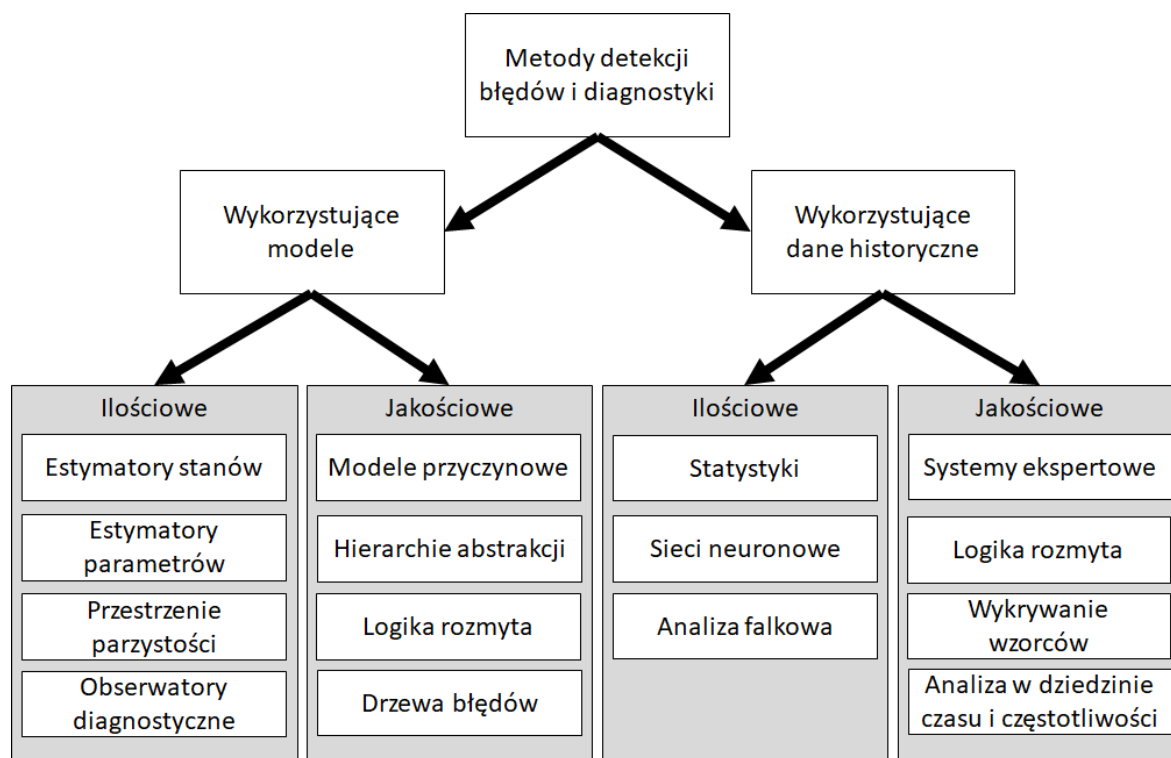
Podejście pierwsze jest bardziej kompleksowe i dające lepsze możliwości detekcji, zwłaszcza znanych i z góry zdefiniowanych typów anomalii, wymaga jednak odpowiedniego zaprojektowania sprzętowego lub ingerencji w istniejącą infrastrukturę, co wiązać się może z koniecznością opracowania dodatkowych dokumentacji dla instalacji, alokacji dodatkowego miejsca produkcyjnego (którego może brakować w okolicach stanowisk produkcyjnych) oraz z poniesieniem dodatkowych nakładów finansowych na budowę lub rozbudowę bazy sprzętowej. Z kolei drugie podejście jest z definicji tańsze i bardziej uniwersalne, wymaga jednak odpowiedniego dostosowania danych procesowych do analiz: większość danych wykorzystywana jest wewnątrz, a systemy produkcyjne nie są dostosowane do ich przechowywania i przetwarzania (np. dane seryjne w których możliwe jest wykrywanie postępujących degradacji zwykle nie są rejestrowane i gromadzone), dodatkowo dane posiadać mogą nieodpowiednią postać dla analiz (np. dla

potrzeb systemów SCADA wykorzystywane są wartości średnie zużycia podczas, gdy wykrywanie anomalii może wymagać wartości chwilowych).

Do detekcji i klasyfikacji błędów różnorakie techniki informatyczne stosowane są już od lat. Najczęściej celem tego typu metod jest szybkie wykrywanie anomalii w procesach produkcyjnych w celu redukcji ryzyka uszkodzenia mienia lub ograniczenia liczby wytwarzanych produktów wybrakowanych, detekcja i analiza postępujących degradacji umożliwiające lepsze zarządzanie serwisowaniem i naprawami maszyn, określanie typów nieprawidłowości w celu polepszenia obsługi błędów produkcyjnych i wsparcia wnioskowania inżynierii procesowej oraz poszukiwanie kroków w procesach, w których występowanie błędów zdarza się najczęściej tak, aby polepszyć jakość lub usprawnić wydajność wytwarzania.

Metody detekcji błędów stanowią pulę rozwiązań znanych w literaturze jako FDD (ang. Fault Detection and Diagnosis). Wielokrotnie zestawiane, klasyfikowane i opisywane były w literaturze, m.in. w [2], [3] czy [4]. Autorzy wskazują najczęściej na cztery typy podejść do rozwiązania tego problemu:

- Z wykorzystaniem modeli ilościowych i procesowych, np. estymatorów stanów, modeli nieliniowych lub równań parametrycznych,
- Z wykorzystaniem modeli jakościowych, np. za pomocą modelowania przyczynowego czy strukturalnymi hierarchiami abstrakcji,
- Z wykorzystaniem danych historycznych i ich analizy jakościowej, np. systemów eksperckich czy logiki rozmytej,
- Z wykorzystaniem danych historycznych i ich analizy ilościowej, np. podejść statystycznych, wykrywania wzorców, w tym klasteryzacji, modeli parametrycznych czy analizy w dziedzinie czasu i częstotliwości.



Rysunek 2.7.1. - Ogólny podział technik wykrywania błędów i diagnostyki urządzeń

Ze względu na różnorakie charaktery analizowanych procesów, w których wykrywane są błędy, jak również stan wiedzy o analizowanym procesie technologicznym czy stosowane techniki pomiarowe do rejestracji sygnałów procesowych, opis dostępnych podejść naukowych i opracowywanych rozwiązań zawężony został do podejść wykorzystujących dane historyczne, w szczególności analizy z wykorzystaniem metod statystycznych, eksploracji danych i analizy w domenie czasu i częstotliwości. Wynika to również z faktu, że badania prowadzone przez autora rozprawy skupione były na wykorzystaniu procesowych danych historycznych rejestrowanych w trakcie pracy stanowisk produkcyjnych, stąd też wykorzystanie tych technik jest najbliższym do metod opracowanych przez autora.

Pierwsza duża pula metod wspierających wykrywanie i klasyfikację anomalii, podobnie jak w przypadku wykrywania wariantów produkcyjnych, realizowana jest z wykorzystaniem analizy przebiegów czasowych. Istnieją liczne opracowania bazujące na analizie szeregów czasowych, ich segmentacji i klasyfikacji dla potrzeb wykrywania anomalii. W pracy [104] zaprezentowane zostało podejście do wykrywania błędów w szeregach danych z wykorzystaniem zmodyfikowanych konwolucyjnych sieci neuronowych. Sieci te rozbudowane zostały o połączenia rezydualne (ang. Residual

Connections), poprawiające jakość działania sieci przy nasyceniu lub stosowaniu większej liczby warstw. Dodatkowo zastosowano zaproponowany przez autorów moduł iniepcji, umożliwiający wykrywanie cech występujących lokalnie i agregujący te cechy w odpowiednich warstwach sieci. Zastosowanie sieci neuronowych przedstawiono również w [105], gdzie przebiegi czasowe sygnałów rejestrowanych przez różne typy czujników śledzących proces produkcji urządzeń mikroelektroniki poddawane były ekstrakcji cech z wykorzystaniem warstwy konwolucyjnej (ang. convolution), funkcji aktywacji (ang. activation functions) i warstwy łączenia (ang. pooling) sieci neuronowych. Następnie z cech budowane były informacje o anomaliach z wykorzystaniem pełnych konwolucyjnych sieci neuronowych, dzięki którym wykrywano i klasyfikowano błędy w napyłaniu cienkowarstwowym na płytkach krzemowych. Detekcja anomalii w oparciu o model zbudowany o zaproponowaną metodę przeprowadzona została dla symulowanego obiektu z prowadzonym procesem chemicznym, w którym wykrywano błędy w pracy instalacji m.in. reaktora i kompresora. Autorzy publikacji [106] opracowali metodę do detekcji anomalii polegającą na dynamicznej selekcji cech (parametrów) mających najbardziej istotny wpływ na proces produkcyjny, budowanie modeli w postaci wzorcowych przebiegów czasowych wyselekcjonowanych cech oraz obliczeniu kontrolnych wartości definiujących nieprawidłowości tak, aby na ich podstawie ustalać przynależność cech bieżących i wykrywać wartości leżące poza ich granicami. Selekcja cech przeprowadzona została z użyciem analizy głównych składowych (PCA) i regresji częściowej najmniejszych kwadratów (ang. Partial Least Squares, PLS), a modelowanie przebiegu z wykorzystaniem wielowarstwowej sieci perceptronowej (ang. Multilayer Perceptron Network, MLP-NN), maszyny wektorów wspierających i lasów losowych. Uruchomienie i walidacja metody przeprowadzona została dla instalacji uzdatniania wody, w której wykrywane były nieprawidłowości w pracy skutkujące problemami w przepływach i jakości wypuszczanej wody. Zastosowanie analizy drgań podczas pracy suwnicy bramowej w branży hutniczej przedstawione zostało w [107]. Autorzy pracy rejestrowali w czasie sygnały z czujników drgań, umożliwiając wyliczenie szeregu wartości diagnostycznych opartych o statystykę, m.in. wartości średnie, wybrane percentyle, kwartyle, mediany czy dominanty oraz składowe w domenie częstotliwości, tj. harmoniczne. Ponieważ liczba potencjalnych cech diagnostycznych była bardzo duża, przeprowadzono selekcję najważniejszych z nich z wykorzystaniem klasyfikatora lasu losowego, a następnie obserwowano trendy zmian w pięciu z nich, przede wszystkim dla

potrzeb detekcji postępujących degradacji w czasie. Przykładami dość powszechnie wykorzystującymi analizę drgań są rozwiązania do detekcji uszkodzeń elementów rotujących, zwłaszcza łożysk. W pracy [108] autorzy analizowali sygnał wibroakustyczny zarejestrowany podczas pracy stanowiska testującego zawieszenie produkowanych wagonów kolejowych. Sygnał zdekomponowany został za pomocą zmodyfikowanej techniki empirycznego trybu dekompozycji (ang. Empirical Mode Decomposition, EMD), nazwanej przez autorów adaptacyjną (AEMD), rozbudowującą krok analizy części oscylacyjnych o dodatkową analizę energii sygnału. Wyznaczone dzięki temu składowe sygnału pierwotnego IMF (ang. Intrinsic Mode Functions), poddawane są dalej przekształceniu do dziedziny częstotliwości, gdzie poprzez analizę porównawczą magnitud dla nominalnej i dzwoniącej częstotliwości pracy łożyska wykrywane są uszkodzenia wynikające ze zużycia pierścieni, np. pęknięcia. W opracowaniu [109] autor wykorzystuje analizę drgań do wykrywania defektów silnika spalinowego na końcowym etapie produkcji, podczas zimnych testów (prowadzonych za pomocą osi elektrycznych wymuszających ruch wału korbowego, bez etapu spalania w komorze). Drgania akustyczne zarejestrowane podczas testów turbiny przekształcone zostały za pomocą transformacji falkowej (ang. Wavelet transformation), a następnie przeanalizowane pod kątem różnic w amplitudzie na wybranych prędkościach obrotowych podczas rozruchu silnika tak, aby wykrywać nieprawidłowe ustawienie wału turbiny. Inne przykłady bazujące na analizie drgań przedstawione zostały w licznych pracach, np. [110], [111] czy [112].

W pracy [113] wykorzystany został klasyfikator lasu losowego do opracowania metody wykrywania wadliwie wytrawianych płytek krzemowych. Wykorzystując analizę spektralną z procesu wytrawiania plazmowego, autorzy wykorzystali miarę podobieństwa obserwacji zbudowaną w oparciu o klasyfikator i wykrywali nieprawidłowo napyłone krzemem płytki oraz dryfy procesu związane z wytrawianiem. Innym przykładem zastosowania klasyfikatora lasu losowego jest [114], w którym został on uzupełniony o algorytmy ewolucyjne selekcji, krzyżowania i mutacji tak, aby ze zbioru danych zebranych z procesu testowania produkowanych napędów asynchronicznych wyłonić najważniejsze cechy do analizy przez klasyfikator, a przez to wypracować najlepsze parametry dla budowy tego klasyfikatora. Efektem działania metody była automatyczna detekcja sześciu najbardziej krytycznych błędów funkcjonalnych powstałych podczas produkcji. Istotnym ograniczeniem obu zaprezentowanych podejść jest fakt, że liczba podziałów i liczba drzew podana musi być z góry, co silnie uzależnia jakość działania

metod od doświadczenia osób strojących algorytm oraz fakt, że umożliwiają operację w oparciu o dane treningowe, co powoduje konieczność przeuczania klasyfikatora przy każdorazowym pojawianiu się nowych typów błędów. W opracowaniu [115] przedstawione zostało wykorzystanie klasyfikacji k -najbliższych sąsiadów i maszyny wektorów wspierających do detekcji błędów w pracy silników asynchronicznych AC. Sygnały akustyczne rejestrowane podczas pracy silników rozbite zostały na składowe częstotliwościowe z wykorzystaniem jednowymiarowej dyskretnej transformaty falkowej (ang. 1-Dimensional Discrete Wavelet Transform, 1D-DWT,). Następnie dokonana została ekstrakcja najważniejszych cech przebiegów z wykorzystaniem jednowymiarowego wzorca binarnego (ang. 1-dimensional Binary Pattern, 1D-BP) po to, by umożliwić klasyfikację stanów poprawności pracy napędu. Z kolei w publikacji [116] autorzy badali możliwości wykrywania i klasyfikacji błędów w działaniu przesyłu energii elektrycznej, wynikających ze zwarć trójfazowych, międzyfazowych i pomiędzy fazami a uziemieniem za pomocą klasyfikatora k -najbliższych sąsiadów. Dane zgromadzone były w formie zapisów o typach błędów, wielkościach fizycznych zarejestrowanych podczas wystąpienia, m.in. rezystancji i mierzonego prądu, oraz lokalizacjach i długościach odcinków transmisyjnych, w których wystąpiły. Zbudowany model posłużył do wykrywania bieżących anomalii w przesyłach i rozbudowywany był w miarę pojawiania się nowych typów błędów. Przykład zastosowania klasyfikatora autoenkodera wariacyjnego (ang. Variational autoencoder, VAE) i połączonej z nim rekurencyjnej sieci neuronowej (ang. Recurrent Neural Network, RNN) zaprezentowany został w [117]. Posłużyły one do detekcji błędów w pracy napędu elektrycznego, takich jak: nieosiowość obciążenia na wale czy obłuzowanie elementów blokujących łożysko, przejawiających się nienaturalnymi drganiami pracującego urządzenia.

Grupa podejść do wykrywania anomalii z wykorzystaniem modeli sygnałowych obejmuje wykorzystanie między innymi analizy głównych składowych (PCA). Przykładem takiego rozwiązania jest [118], w którym autorzy wykorzystali ją do selekcji krytycznych parametrów procesowych wpływających na błędy w produkcji mikroelektroniki oraz do określenia kroków w procesie, w którym te parametry wykazują największe zmiany. Wybór najistotniejszych parametrów wspierany jest klasyfikacją z wykorzystaniem maszyny wektorów podpierających oraz algorytmu AdaBoost. Po ich selekcji, zastosowany został algorytm drzew decyzyjnych, w którym czas realizacji procesu przekształcony do postaci dyskretnej wykorzystywany jest jako atrybut dzielenia drzewa.

Ocena podziału ze względu na największą wartość wzmocnienia informacyjnego umożliwiła wskazanie na stempel czasowy, w którym wystąpiła zmiana badanego parametru, a tym samym wskazanie kroku w procesie, w którym potencjalnie wystąpiła anomalia. Innym przykładem jest rozwiązanie opisane w [119], w którym autorzy zastosowali analizę głównych składowych (PCA) do wykrywania sytuacji anomalnych w pracy pieca do pirolizy. Parametry procesowe związane z wykrywanymi błędami poddawane były następnie porównaniu z T-kwadratową statystyką Hotellinga (T^2), a w przypadku przekroczenia dopuszczalnych wartości i zbyt długiego czasu trwania ustalane jako generujące anomalię. Następnie model grafów skierowanych oparty o dodatkową metodę PCA umożliwił reprezentację powiązań pomiędzy zmiennymi w procesie i wskazanie na te jego kroki, w których wystąpiła anomalia. Autorzy opracowania [120] przedstawili metodę do analizy metalowych komponentów łączonych w technologii walcowania na gorąco. Wykorzystali w niej zaproponowaną przez siebie technikę klasyfikacji danych wieloetykietowych opartą m.in. o aproksymację Nyströma (ang. Nyström approximation) i metodę analizy dyskryminacyjnej Fischera (ang. Fisher Discrimination Analysis, FDA). Dzięki niej możliwe było wykrywanie i klasyfikowanie anomalii powstałych na skutek doboru niewłaściwych nastaw procesu, m.in. sił oddziałujących na walcowane elementy i prędkości przenoszenia, skutkujących defektami w wytwarzanych produktach końcowych.

Do detekcji błędów w pracy instalacji przemysłowych stosowane są również techniki klasteryzacji. Choć rozwiązania te nie są bardzo popularne, stosować je można zarówno w postaci głównego algorytmu analitycznego do detekcji anomalii jak również jako techniki wspomagające budowę modeli wzorcowych pracy instalacji. Właściwości wybranych algorytmów klasteryzacyjnych, m.in. zdolność do wyznaczania wartości odstających obserwacji przez algorytm k -Means w postaci badania odległości do centroidów klastrów, wykorzystana została w pracy [121]. W przedstawionej metodzie zastosowano podwójny mechanizm klasteryzacji, który w pierwszym kroku wyznacza klastry zgodnie z wyznaczoną ich liczbą, a następnie redukuje lub zwiększa ich liczbę przy jednoczesnym ponownym dopasowaniu w przypadku, gdy występują obserwacje znacząco oddalone od wyliczonych centroidów. W pracy [122] zaproponowane zostało podejście do wykrywania błędów w pracy osi numerycznych maszyn, w szczególności uszkodzeń łożysk, poprzez selekcję z przebiegu drgań informacji o błędach w oparciu o transformację falkową i klasteryzację metodą propagacji podobieństwa (ang. Affinity Propagation).

Podobne zagadnienia badane były i zostały opisane przez autorów pracy [123], w której algorytm klasyfikacji błędów oparty został o Ukryty Model Markova (ang. Hidden Markov Model) i aglomeracyjną klasteryzację rojem (ang. Swarm Clustering). Autorzy publikacji [124] zastosowali aglomeracyjną metodę klasteryzacji DBSCAN do wykrywania błędów w pracy sprzężarek śrubowych tak, aby umożliwić naukę modeli opisujących pracę tych urządzeń z wykorzystaniem analizy PCA. Oprócz sygnałów wibroakustycznych rejestrowane było również zużycie energii, wspomagające pracę mechanizmów analitycznych. W przedstawionych powyżej przykładach do nauki modeli lub zbiorów referencyjnych konieczne było stosowanie wysokiej jakości niezaburzonych danych produkcyjnych oraz wiedzy na temat typów anomalii do wykrywania. Ograniczenie i trudności związane z akwizycją tego typu danych, wynikające z dużej liczby elementów wykonawczych, mnogości sygnałów czy sposobów prowadzenia procesów zauważono w [125]. Z tego powodu połączono klasyfikator autoenkodera do wyodrębnienia cech z sygnałów drgań z pracy napędów z nauką metryki odległości (ang. Distance Metric Learning) i algorytmu *k*-Means w architekturze sieci neuronowych. Umożliwiło to wykrywanie błędów bez odgórnej wiedzy na temat ich liczby czy właściwości.

Wykrywanie anomalii z wykorzystaniem technik eksploracji danych jest skomplikowanym zadaniem, ale przy zastosowaniu odpowiednich metod może prowadzić do interesujących i wartościowych rezultatów, wpływających nie tylko na budowanie wiedzy o procesach, lecz również do wsparcia utrzymania instalacji przemysłowych, m.in. przez wczesne wykrywanie awarii lub wykrywanie postępującej degradacji urządzeń w czasie. Choć zagadnienia te, zwane predykcyjnym i prewencyjnym utrzymaniem zyskują na popularności, to w przeważającej większości w opracowaniach literaturowych koncentrują się na analizach znanej i szeroko opisywanej w literaturze pracy napędów i składowych osi numerycznych. Dodatkowo najczęściej wymagają stosowania zewnętrznych urządzeń pomiarowych dedykowanych do rejestracji sygnałów poddawanych analizie, nierzadko ograniczając w ten sposób wybór dostępnych metod analitycznych. W opracowaniach naukowych rzadko opisywane są podejścia do wykrywania anomalii w pracy całych instalacji produkcyjnych, w szczególności w produkcji dyskretniej. W związku z tym prace autora prowadzone były na instalacjach obsługujących procesy dyskretnie, w których do sterowania wykorzystywano sterowniki swobodnie programowalne. Dzięki temu możliwy był bezpośredni dostęp do sygnałów procesowych, który oprócz stosowania zewnętrznych mierników zużycia energii, nie

wymagał wykorzystywania dedykowanych układów pomiarowych. Budowanie informacji diagnostycznych prowadzone więc było bez ingerencji w technologię oraz w oparciu o sygnały sterujące stanowiska, umożliwiając w ten sposób generalizację rozpatrywanych problemów badawczych i zapewniając skalowalność metod niezależnie od technologii i dziedziny.

3. BADANIA NAD ZASTOSOWANIEM MODELI INFORMACYJNYCH DO WSPARCIA UTRZYMANIA INSTALACJI PRZEMYSŁOWYCH

Nowoczesne systemy realizacji produkcji nie tylko powinny dostarczać kompleksowej informacji o stanie i jakości realizowanych procesów, ale dodatkowo powinny budować i rozszerzać wiedzę na temat produkcji. Jednym ze sposobów uzupełniania tej wiedzy jest połączenie danych procesowych, generowanych na poziomie urządzeń wykonawczych i kontrolerów procesu oraz wyników przetwarzania tych danych z wiedzą o instalacji produkcyjnej. Dzięki takiemu powiązaniu możliwe jest selekcjonowanie danych niezbędnych do analiz oraz wiązanie typów błędów produkcyjnych z miejscami i obszarami produkcji, w których występują, wskazywanie na potencjalne warianty produktów lub ich składowe mogące wpływać negatywnie na jakość realizowanego procesu, a w ujęciu holistycznym wspomagać procesy utrzymania urządzeń. W rozdziale 2.2. opisane zostały możliwości modelowania fragmentów instalacji tak, aby wykorzystywać wiedzę do zarządzania procesem produkcyjnym. Te same modele mogą zostać wykorzystane do wyboru odpowiednich wartości procesowych do budowania wiedzy o jakości i efektywności produkcji, w szczególności przez automatyczną klasyfikację cykli produkcyjnych i wykrywanie anomalii w pracy instalacji przejawiających się odchyleniami od pracy wzorcowej. Rezultaty z badań opisanych w niniejszym rozdziale prezentują rozwiązanie problemu opisanego w rozdziale 1., polegającego na zastosowaniu modeli informacyjnych do wiązania danych o procesie z danymi technologicznymi i stanowią wprowadzenie do metod budowy wektorów cech wykorzystywanych w zadaniach analitycznych, opisanych w rozdziałach 4 i 5.

3.1. Model informacyjny technologii instalacji produkcyjnej

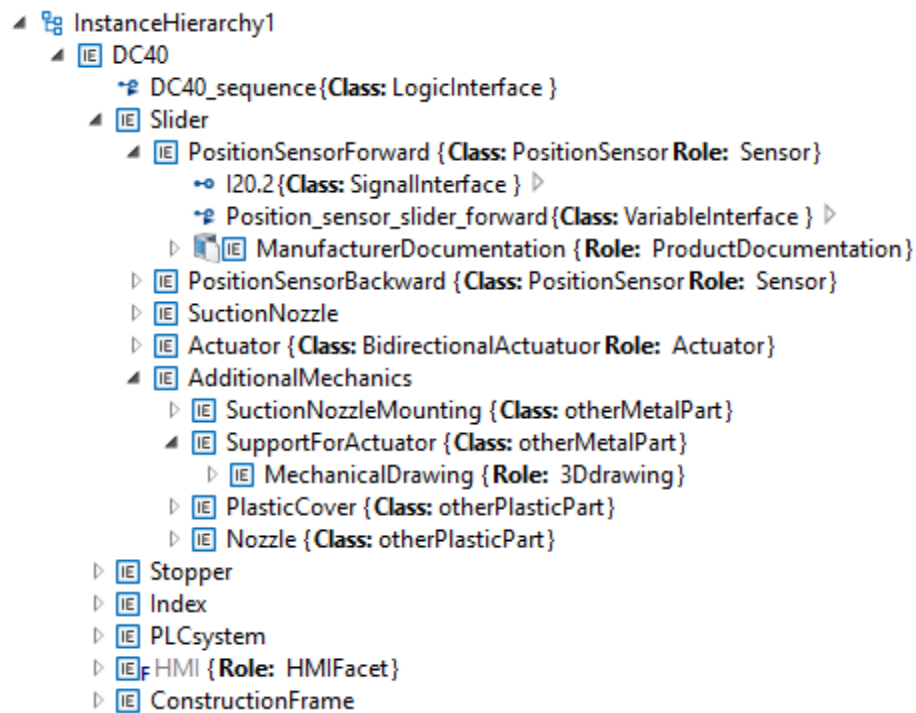
W rozdziale 2.2. przedstawiona została możliwość zamodelowania procesu produkcyjnego z wykorzystaniem standardu ANSI/ISA-95. Jednocześnie opisane zostały powody dla których współczesne modele informacyjne dedykowane dla rozwiązań przemysłowych nie pasują do wymagań produkcji masowo-dostosowanej. Z tego powodu konieczne jest opracowanie nowego podejścia wiążącego dane przemysłowe z wiedzą o technologii, które będzie zgodne z architekturą RAMI4.0, opisaną w rozdziale 2.1. Dla potrzeb zbadania tej możliwości, autor rozprawy przeprowadził badania nad utworzeniem

cyfrowego bliźniaka instalacji, który oparty o nowoczesny model informacyjny umożliwi powiązanie danych procesowych ze składowymi fizycznymi badanej instalacji. Dzięki temu możliwe będzie wykorzystanie wiedzy o wykrywanych anomaliach w pracy urządzeń wykonawczych dla potrzeb wsparcia utrzymania instalacji produkcyjnej. Cyfrowy bliźniak zbudowany został jako odzwierciedlenie rzeczywistej struktury urządzeń przemysłowych znajdujących się na stanowisku produkcyjnym, wraz z zachowaniem ich parametrów, właściwości, relacji, które wiążą je z ich składowymi i które wiążą poszczególne komponenty.

Modelowanie technologii produkcyjnej przeprowadzone zostało z wykorzystaniem języka Automation Markup Language. Język ten umożliwia opisanie fizycznej struktury instalacji w postaci hierarchii instancji, gdzie poszczególne urządzenia wraz z ich atrybutami związane są w grupy. Model wspiera podejście obiektowe, tj. możliwe jest definiowanie klas urządzeń i grup, wykorzystywanych następnie w postaci instancji. Dodatkowo, umożliwia modelowanie korelacji pomiędzy obiektami, np. połączeń fizycznych (interfejsów komunikacyjnych, elektrycznych i mechanicznych), logicznych (zależności pomiędzy elementami instalacji) i funkcjonalnych (zasobów związanych z danymi elementami, np. dokumentacjami). Oprócz podstawowych elementów infrastruktury produkcyjnej możliwe jest również modelowanie elementów pomocniczych stanowiących wyposażenie stanowiska, np. szaf elektrycznych, okablowania, stacji przygotowania powietrza, systemów bezpieczeństwa czy elementów HMI. Język ten został wykorzystany ze względu na rozszerzającą się popularność w środowiskach przemysłowych oraz prostotę wykorzystania opartą o format XML (ang. Extensible Markup Language).

Badania nad zastosowaniem modeli danych do powiązania z danymi procesowymi prowadzone były z wykorzystaniem eksperymentalnego stanowiska produkcyjnego DC40, opisanego w załączniku 1. Fragment utworzonej hierarchii instancji dla stanowiska DC40 przedstawiony został na rysunku 3.1.1. Hierarchia przedstawiona jest w postaci drzewa, a poszczególne poziomy zagnieżdżeń przedstawiają zależności pomiędzy elementami struktury. Najwyższym jej poziomem jest cała instalacja produkcyjna, a więc kompletne stanowisko DC40. Poniższe segmenty tworzone są przez zgrupowania funkcjonalne urządzeń (np. urządzenia wykonawcze wraz z towarzyszącym im oczujnikowaniem) oraz pojedyncze urządzenia wykonawcze. Niektóre z zamodelowanych obiektów posiadają przypisane role, np. elementy składowe mechaniki stanowią elementy konstrukcyjne

stanowiska, a sensory pełnią rolę systemu monitorującego i odczytania stacji. Interfejsy poszczególnych obiektów widoczne są jako powiązania z dokumentami zewnętrznymi, np. zasobami technologicznymi czy plikami projektowymi, oraz jako połączenia wewnętrzne, wskazujące na logiczne i funkcjonalne związki z innymi obiektami. Uzupełnieniem dla całości hierarchii są atrybuty, nadawane na poziomie definicji poszczególnych klas, umożliwiające wyspecyfikowanie właściwości i parametrów poszczególnych obiektów.



Rysunek 3.1.1. – fragment hierarchii instancji w języku AutomationML przygotowany dla stanowiska DC40

Rozwinięcie modelu w postaci szczegółowych wpisów w formacie XML przedstawione jest na rysunku 3.1.2.

Wszystkie obiekty zamodelowane w języku AutomationML reprezentowane są przez elementy wewnętrzne (ang. *Internal Element*). Najwyższym elementem hierarchii jest instancja o nazwie DC40. Poniżej znajdują się grupy urządzeń nazwane od głównych urządzeń wykonawczych stanowiska, np. wypychacz (ang. Slider) czy podnośnik (ang. Index). Grupy składają się z urządzeń wykonawczych, tj. siłowników lub zaworów pneumatycznych oraz urządzeń sygnalizujących stany, np. czujników położenia tłoka wypychacza w przodzie (ang. Position sensor slider forward). Każdy z elementów posiadać

może różne atrybuty, takie jak określenie producenta (ang. *Manufacturer*) lub identyfikator (ang. *Identifier*). Jednocześnie elementy posiadać mogą interfejsy zewnętrzne (ang. *External Interface*), definiujące powiązanie z elementami zewnętrznymi, np. podłączenie do konkretnego wyjścia sterującego PLC (np. Q20.2) czy zasobu w postaci dokumentacji (np. ścieżki do dokumentu zewnętrznego).

```

- <InstanceHierarchy Name="InstanceHierarchy1">
  <Version>1.0.0</Version>
  - <InternalElement Name="DC40" ID="d3496a98-e9b0-4342-b3d5-6436ec640a2b">
    - <ExternalInterface Name="DC40_sequence" ID="80521a0d-bc4e-4ade-a1c2-f609812c3f9a" RefBaseClassPath="AutomationMLIN"
      - <Attribute Name="refURI" AttributeDataType="xs:anyURI">
        <Value>file:///Z:/AutoUniMo/DC40/PLC/DC40 - simplified.xml# DC40_sequence</Value>
      </Attribute>
    </ExternalInterface>
  - <InternalElement Name="Slider" ID="f3f7290f-b2e5-453d-a0f4-8aab6348574f">
    - <InternalElement Name="PositionSensorForward" ID="d35ad262-446d-4e6d-9032-dcb48c08c966" RefBaseSystemUnitPath="
      - <Revision>
        <RevisionDate>2016-11-04T11:19:11.7518095</RevisionDate>
        <AuthorName>Please enter your name</AuthorName>
      </Revision>
      <AdditionalInformation/>
      <Attribute Name="Attribute1"/>
    - <ExternalInterface Name="I20.2" ID="23bcf7c4-c1db-4d14-b774-78640435a5cb" RefBaseClassPath="AutomationMLInter
      - <Revision>
        <RevisionDate>2016-11-04T11:26:03.6798095</RevisionDate>
        <AuthorName>Please enter your name</AuthorName>
      </Revision>
      <AdditionalInformation/>
    </ExternalInterface>
  - <ExternalInterface Name="Position_sensor_slider_forward" ID="dc1cca2c-c110-41f9-b4c2-6e3fcd194004"
    RefBaseClassPath="AutomationMLInterfaceClassLib/AutomationMLBaseInterface/ExternalDataConnector/PLCopenXML
      - <Revision>
        <RevisionDate>2016-11-04T11:17:28.5288095</RevisionDate>
        <AuthorName>Please enter your name</AuthorName>
      </Revision>
      <AdditionalInformation/>
      - <Attribute Name="refURI" AttributeDataType="xs:anyURI">
        <Value>file:///Z:/AutoUniMo/DC40/PLC/DC40 - simplified.xml# Position_sensor_slider_forward</Value>
      </Attribute>
    </ExternalInterface>
  - <InternalElement Name="ManufacturerDocumentation" ID="c2c0d1b5-bde8-4567-acb0-f79a93b6fbd4">
    - <Revision>
      <RevisionDate>2016-11-04T11:24:35.5118095</RevisionDate>
      <AuthorName>Please enter your name</AuthorName>
    </Revision>
    <AdditionalInformation/>
    <Attribute Name="Attribute1">
      <RefSemantic CorrespondingAttributePath=""/>
    </Attribute>
  - <ExternalInterface Name="DocumentLink" ID="96b4025f-20f9-4cf-a232-286bd92eb377" RefBaseClassPath="Automa
    - <Attribute Name="MIMEType" AttributeDataType="xs:string">
      <Value>/pdf</Value>
    </Attribute>
    - <Attribute Name="refURI" AttributeDataType="xs:anyURI">
      <Value>file:///Z:/AutoUniMo/DC40/Manufacturer documentations#PositionSensor slider stop prestop dal
    </Attribute>
  </InternalElement>
</InstanceHierarchy>

```

Rysunek 3.1.2. – fragment hierarchii instancji w formacie XML przygotowany dla stanowiska DC40

Elementy posiadają dodatkowo role (ang. *Role Requirements, Supported Roles*) umożliwiające klasyfikację ze względu na pełnione funkcje, np. odczytywanie dla urządzeń monitorujących (ang. *Indication Sensor*) lub urządzenie kontrolne dla jednostek sterujących (ang. *Control Device*). Każdy z elementów posiada unikalny identyfikator zapisany w polu ID. Identyfikatory mogą zostać wykorzystane do przetwarzania modelu, np. do poszukiwania konkretnego elementu lub jako konektor do podłączenia systemów zewnętrznych.

Model informacyjny technologii wykonania stanowiska produkcyjnego stanowi pierwszy element niezbędny do powiązania rezultatów z przetwarzania analitycznego danych procesowych i wiązania ich wiedzą technologiczną dla potrzeb wsparcia

utrzymania instalacji przemysłowej. Opisany za pomocą języka AML, wykorzystany został w dalszych częściach badań.

3.2. Model informacyjny wzorców technologicznych

Informacje o wariantach produktów, które mają zostać wytworzone przez ciąg technologiczny, a przez to na pojedynczych stanowiskach wykonawczych, typowo przygotowywane są i zarządzane przez nadrzędne systemy produkcyjne, np. klasy MES. Układy wykonawcze na poziomie automatyki procesowej często nie posiadają tej informacji, a jedynie wykonują zadania produkcyjne zgodnie z dostarczonymi odgórnie nastawami, przekazywanymi na przykład za pośrednictwem znaczników RFID, identyfikatorów zadań lub kodów kreskowych.

Jednym ze sposobów formalnego opisu odpowiedzi procesowej na realizowany wariant produkcyjny jest zaproponowany przez autora wzorzec technologiczny. Może on zostać zdefiniowany jako wektor zawierający cechy charakterystyczne dla danego wariantu produkcyjnego, które związane są ze specyfiką jego wykonywania w trakcie pracy instalacji. O ile wariant produkcyjny definiowany i stosowany jest przez personel zakładowy i to na jego podstawie wprowadzane są odpowiednie nastawy do urządzeń wykonawczych, o tyle wzorzec technologiczny jest faktycznym odzwierciedleniem pracy instalacji przy realizowanym wariantcie, które jest reprezentowane przez zestaw wybranych wartości procesowych, definiujących cykl produkcyjny, tj. czasookres pracy instalacji, w trakcie którego wytwarzany jest pojedynczy produkt.

Rozmiar i indywidualne cechy wzorca technologicznego mogą się różnić ze względu na złożoność instalacji i zastosowane technologie, jednak dla produkcji dyskretnej możliwe jest zdefiniowanie jego uogólnionej formy zawierającej składowe:

- Całkowity czas trwania cyklu produkcyjnego, definiowany jako czas produkcji danego wariantu produkcyjnego na stanowisku,
- Informacje o wykorzystywanych narzędziach (urządzeniach wykonawczych) w trakcie realizacji cyklu produkcyjnego,
- Czasy aktywności poszczególnych narzędzi w trakcie realizacji cyklu produkcyjnego,

- Całkowite zużycie zasobów produkcyjnych niezbędnych do przeprowadzenia cyklu produkcyjnego, np. jednostkowych materiałów produkcyjnych lub surowców technologicznych, np. olejów w litrach,
- Zużycie zasobów produkcyjnych przez pojedyncze urządzenia wykonawcze lub ich grupy,
- Całkowite zużycie energii zasilającej do przeprowadzenia pełnego cyklu produkcyjnego, np. energii elektrycznej w watogodzinach,
- Zużycie energii zasilającej pojedyncze urządzenia wykonawcze lub ich grupy.

Cechy wektora technologicznego wskazują na wielkości i zapotrzebowanie surowcowe niezbędne do produkcji danego wariantu produktu. Z tego powodu traktowane mogą być jako zasoby wymagane do prowadzenia procesu, a więc możliwe jest ich zamodelowanie w postaci informacji materiałowych (ang. *Material Information*) standardu ANSI/ISA-95. Dzięki niemu możliwe jest opisanie specyfiki każdego z analizowanych wzorców technologicznych w postaci zestawu wartości, odpowiadających wartościom cech referencyjnych wektora V_{MC} . Fragment modelu materiałowego opisującego wzorce technologiczne przedstawiony został na rysunku 3.2.1.

Zgodnie ze standardem ANSI/ISA-95, model materiałowy obejmuje cztery poziomy informacji: klasa materiału (ang. *MaterialClass*), zawierająca ogólne informacje o właściwościach materiałów, będące składowymi wspólnymi pomiędzy klasami materiałów, definicja materiału (ang. *MaterialDefinition*), stosowana do szczegółowej definicji materiału, partia materiału (ang. *MaterialLot*), wykorzystywana do opisu wsadu materiałowego potrzebnego do produkcji w segmencie procesu oraz podpartia materiału (ang. *MaterialSublot*), umożliwiająca szczegółowy opis składowych partii materiałowej. W przedstawionym przykładzie model materiałowy obejmuje zestaw sygnałów charakteryzujących cykle produkcyjne stanowiska DC40, m.in. całkowite zużycie sprężonego powietrza czy czas pracy nad pojedynczym wariantem. W szczególności zawierać może informacje o wzorcach technologicznych, tj. referencyjnych wartościach zużycia i czasów podczas pracy stanowiska nad poszczególnymi wariantami produkcji, podawanych zarówno przez eksperta dziedzinowego (np. inżyniera produkcji lub technologa) jak i z poziomu systemu do automatycznego wykrywania wzorców technologicznych, przedstawionego w rozdziale 4. Wzorce odpowiadają wariantom produktów wytwarzanych za pomocą instalacji i zawierają szczegółowe informacje na temat wartości średnich, minimalnych i maksymalnych zużycia powietrza oraz czasu

aktywności poszczególnych urządzeń wykonawczych. Ponieważ stanowisko DC40 nie zużywa dodatkowych komponentów produkcyjnych (np. części), model materiałowy obejmuje jedynie czas produkcji i zużycie energii. W klasycznych przypadkach produkcyjnych model materiałowy obejmowałby ich zużycie w zależności do produkowanego typu.

Tag	Value
ID	Technological template
Description	Definition of technological tem...
MaterialClassProperty	
MaterialClassProperty	
ID	Average_air_consumption
Description	Average consumption of air in ...
Value	N
Unit	litres
MaterialClassProperty	
MaterialClassProperty	
MaterialClassProperty	
MaterialClassProperty	
MaterialClassProperty	
MaterialDefinition	
ID	Template_1
Description	Details of technological templat...
Hierarchy	
MaterialDefinitionProperty	
ID	Average_air_consumption
Description	Total consumption of air in tem...
Value	7,3317
Unit	litres
MaterialDefinitionProperty	
MaterialDefinitionProperty	
MaterialDefinitionProperty	
ID	Average_activation_time
Description	Average activation time of dev...
Value	1.784
Unit	milliseconds
MaterialDefinitionProperty	
ID	Index_up
Description	
MaterialDefinitionProperty	
ID	Address
Value	Q20.0
MaterialDefinitionProperty	
MaterialDefinitionProperty	

Rysunek 3.2.1. – fragment modelu materiałowego opisującego wzorce technologiczne stanowiska DC40

Choć każda z opisywanych właściwości materiałowych posiada swój identyfikator, więcej informacji na ich temat opisuje opcjonalne pole opisu (ang. *Description*). Dodatkowymi polami są wartość (ang. *Value*) oraz jednostka (ang. *Unit*), definiujące odpowiednio: zużycie powietrza oraz czas aktywności, wyrażone dla trzech typów klas i definicji: wartości średniej, minimalnej i maksymalnej. Właściwości modeli

materiałowych mogą być zagnieżdżone, a przez to model może być skomplikowany, zwłaszcza dla produkcji charakteryzujących się dużą liczbą komponentów montażowych wymaganych do kompletacji i wytwarzania.

Dla opisywanego przypadku poszczególne definicje czasu aktywacji zawierają dodatkowe informacje o urządzeniach wykonawczych, w szczególności o ich adresach wyjściowych systemu sterującego PLC.

Dzięki modelowi informacyjnemu materiałów możliwe będzie wskazanie na aktualnie produkowany wariant produkcyjny, a dzięki temu określenie czy poszczególne składowe technologii działają prawidłowo. Do powiązania modelu materiałowego z modelem technologicznym instalacji wykorzystane zostaną wartości w polach tych modeli, m.in. pole z adresem PLC urządzeń wykonawczych i pola z aktualnymi wartościami zużycia i czasów aktywacji poszczególnych urządzeń.

3.3. Integracja danych procesowych z wiedzą technologiczną

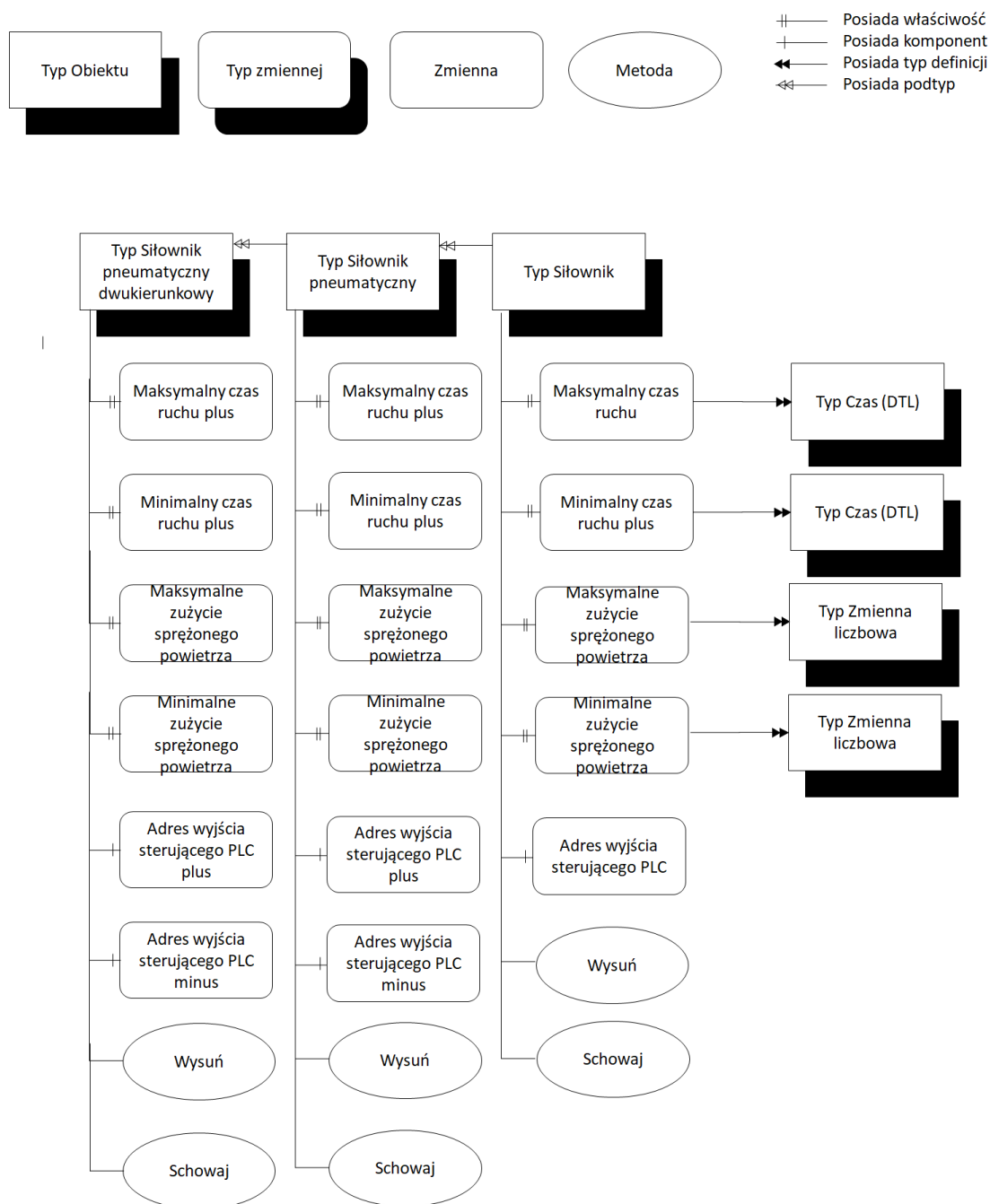
Celem prac badawczych opisanych w niniejszej rozprawie jest analiza efektywności i jakości pracy instalacji przemysłowych poprzez detekcję anomalii w cyklach produkcyjnych oraz automatyczną klasyfikację wariantów produkcyjnych w oparciu o wzorce technologiczne. Informacje te nie są możliwe do bezpośredniego ujęcia w modelu informacyjnym ANSI/ISA-95, ale dzięki zastosowaniu odpowiednich mechanizmów wiążących dane o technologii z miejscami wykrytych anomalii, możliwe jest wskazanie na źródła wystąpienia tych problemów, a przez to uzupełnienie luki w modelu informacyjnym ANSI/ISA-95 i spełnienie założeń architektury RAMI4.0. Jedną z możliwości takiej integracji zapewnia standard komunikacyjny OPC UA, opisany w rozdziale 2.3. Właściwości przestrzeni adresowej OPC UA umożliwiają nie tylko zamodelowanie struktury fizycznej instalacji ale również powiązanie danych procesowych związanych z urządzeniem dla którego wykryto problemy z elementem hierarchii opracowanym w języku AutomationML za pomocą wybranych wartości i zmiennych. Dzięki zastosowaniu takiego podejścia możliwe jest nie tylko utworzenie cyfrowego bliźniaka fizycznej instalacji produkcyjnej ale również przepływu danych, w tym diagnostycznych, związanych z pracą urządzeń wykonawczych. Jednocześnie jest to zgodne z oboma osiami poziomymi architektury RAMI4.0, stanowiącej modelową architekturę dla nowoczesnych systemów wytwarzania, w której odzwierciedlona jest nie tylko hierarchia instalacji fizycznej ale również cykl życia i strumień wartości związane z instalacją.

Zgodnie z modelem interfejsu OPC UA, złożona struktura informacji o instalacji przemysłowej może zostać opisana przez mniejsze składowe powiązane ze sobą odpowiednimi referencjami, wskazujące na typ wiążących je zależności. Możliwe jest opisanie zarówno obiektów prostych, jak i złożonych, składających się z wielu różnych podtypów czy właściwości. Podejście to odpowiada modelowaniu zorientowanemu obiektowo, w którym grupy obiektów o podobnych cechach tworzone są przez odpowiednie typy, a poszczególne instancje odpowiadają pojedynczym, rzeczywistym elementom infrastruktury. Jednocześnie, ponieważ model ten powinien stanowić warstwę pośrednią w komunikacji pomiędzy procesem i potencjalnymi systemami zarządzania wiedzą o technologii czy utrzymaniu instalacji przemysłowej, jego struktura powinna zostać oparta o model ANSI/ISA-95. Przestrzeń adresowa OPC UA stanowić więc będzie zbiór różnych typów definicji i instancji, odpowiadających rzeczywistym elementom wykonawczym, a których hierarchia odwzorowana jest w postaci modelu segmentu procesowego zgodnego z ANSI/ISA-95.

Przygotowanie przestrzeni adresowej OPC UA przeprowadzone zostało zgodnie z zasadami zaproponowanymi przez OPC Foundation ISA-95 Working Group, polegającymi na traktowaniu wyposażenia (ang. *Equipment*) i klas wyposażenia (ang. *Equipment Class*) wskazanych przez ANSI/ISA-95 jako obiektów OPC UA, a właściwości sprzętu (ang. *Property*) jako zmiennych OPC UA. Podobne założenia powinny zostać podjęte przy modelowaniu innych składowych ANSI/ISA-95, np. personelu (ang. *Personnel*) czy zasobów (ang. *Material*).

Nadrzędną składową, w oparciu o którą opracowano przestrzeń adresową OPC UA był segment procesu, przedstawiony w rozdziale 2.2. Każdy z segmentów procesu opisany został przez dwa komponenty: modele informacyjne wykorzystywanego wyposażenia (ang. *Equipment Information*) i zużywanych zasobów (ang. *Material Information*). W skład wyposażenia weszły główne komponenty stanowiska DC40 wraz z oprzyrządowaniem, a jako zużywane materiały traktowano czasy aktywacji urządzeń oraz całkowite zużycie sprężonego powietrza. Utworzona struktura zbudowana została w postaci zorientowanej obiektowo oraz rekursywnej, gdzie poszczególne komponenty posiadają swoją definicję typu w postaci klasy w modelu ANSI/ISA-95, a poszczególne składowe podrzędne mogą stanowić zbiór formujący składową nadrzędną. Przykład

fragmentu tego modelu dla siłownika z wybranymi właściwościami i metodami przedstawiony został na rysunku 3.3.1.



Rysunek 3.3.1. – fragment modelu instalacji DC40 opartego o definicję typów

Nadrzędnym obiektem jest typ Siłownik, stanowiący klasę wyposażenia (ang. *Equipment Class*), definiującą najbardziej ogólne właściwości tego typu urządzeń. Do urządzeń tej klasy należą zarówno siłowniki pneumatyczne jedno- i dwukierunkowe, z których zbudowane jest stanowisko DC40, jak również elektrosiłowniki lub jakiegokolwiek inne urządzenia charakteryzujące się posuwem tłoka i technologicznie zbliżone do tego

sprzętu. Typ ten posiada swoją ogólną definicję zawierającą wybrane parametry odpowiadające właściwościom (ang. *Property*) oraz komponenty (ang. *Equipment*) i metody. Ogólnie zdefiniowane urządzenia posiadać mogą podtypy czyli uszczegółowione wersje różnych wariantów technologicznych urządzeń wykonawczych, na przykład *Siłownik pneumatyczny*, będący szczególnym typem siłownika, oraz *Siłownik pneumatyczny dwukierunkowy*, będący szczególnym typem *Siłownika pneumatycznego*. Różne podtypy mogą posiadać szczegółowe i różniące się od siebie składowe, np. związane z technologią wykonania urządzenia wartości referencyjne czy komponenty.

W rozdziale 5. przedstawiona została metoda automatycznego wykrywania anomalii w pracy instalacji. Zakłada ona wykorzystanie wektorów opisujących cykle produkcyjne V_{MC} , których poszczególne cechy związane są z parametrami pracy instalacji takimi jak całkowite czasy aktywności urządzeń wykonawczych i całkowite zużycie powietrza podczas pracy stacji. Powiązanie faktu wystąpienia anomalii z konkretnym urządzeniem wykonawczym leży poza klasycznymi możliwościami modelu ANSI/ISA-95, jednak dzięki zastosowaniu modelu technologicznego opartego o AML i dostępowi do niego z wykorzystaniem przestrzeni adresowej OPC UA możliwe jest powiązanie wszelkich nieprawidłowości ze wszystkimi zasobami technologicznymi związanymi z awaryjnym urządzeniem wykonawczym (np. dokumentacjami, plikami projektowymi lub sygnałami programowymi). Powiązanie odbywać się może za pomocą kilku pól:

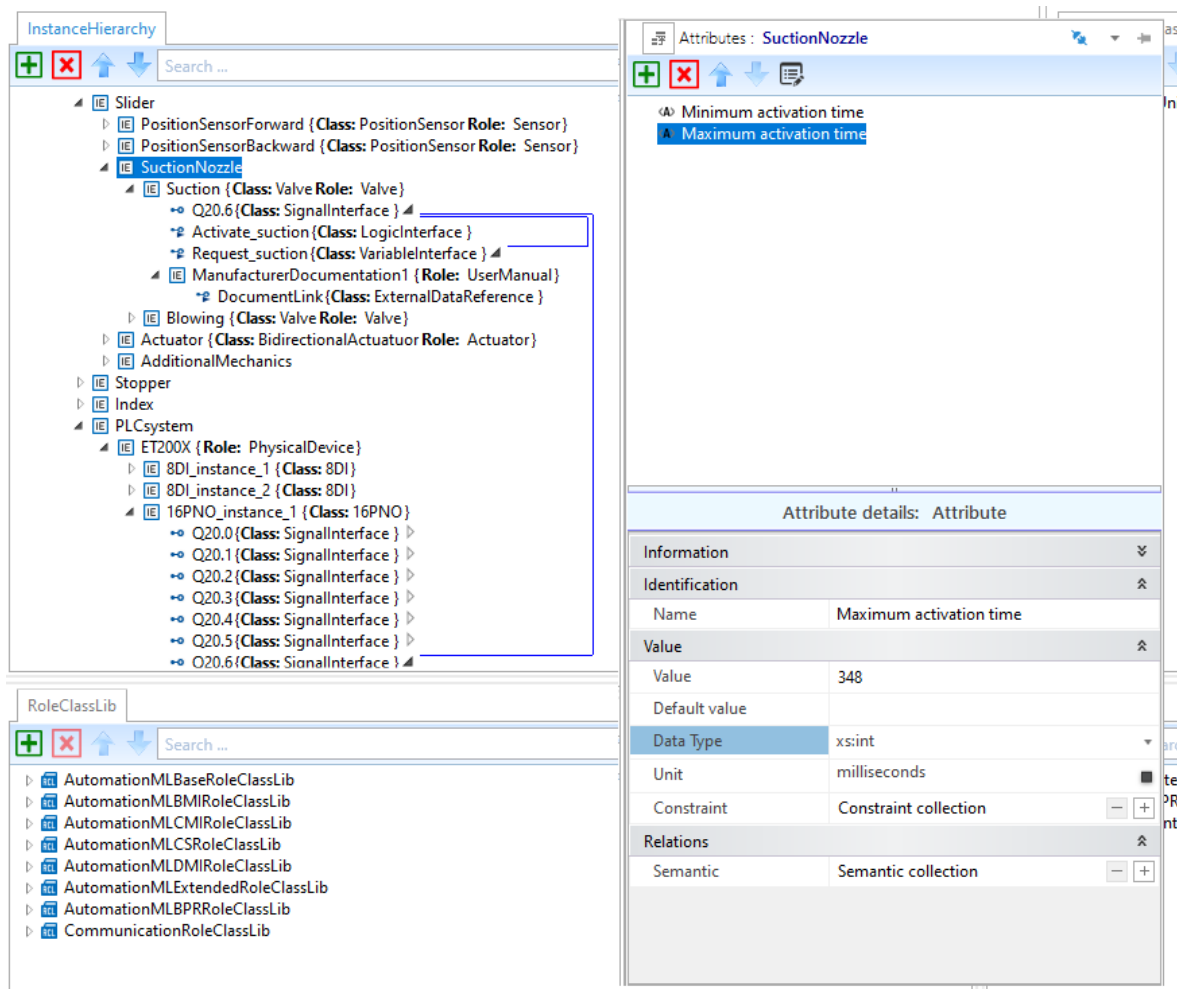
- Globalnego identyfikatora urządzenia – stosowanego dla przypadku rejestrowania ID urządzenia wykonawczego z rezultatem procesu produkcyjnego, np. automatycznego testera funkcjonalnego pozostawiającego swój cyfrowy podpis produkcyjny w postaci identyfikatora, wykorzystanego jednocześnie do identyfikacji węzła OPC UA i elementu wewnętrznego AML,
- Adresu wejścia lub wyjścia sterującego PLC - wskazującego na konkretne urządzenie wykonawcze realizujące czynność procesową, zapisanego jako zmienna OPC UA i atrybut AML,
- Wartości mierzonej porównywanej do wartości referencyjnych – odpowiadającej wynikowi przeprowadzonej operacji procesowej lub wskazującej na jakość tej czynności (np. czas wykonywania operacji lub sumarycznego zużycia energii), zapisanej jako zmienna OPC UA i atrybut AML.

Dla potrzeb badań autor wybrał powiązanie przez adres wyjścia sterującego PLC w połączeniu przez wartości mierzone całkowitego zużycia powietrza w poszczególnych cyklach produkcyjnych. Wynika to z faktu rejestracji danych procesowych za pomocą monitorowania stanu tych właśnie sygnałów PLC i wyznaczania za ich pomocą całkowitego czasu aktywności poszczególnych urządzeń wykonawczych. Jednocześnie wartości referencyjne zużyc i czasów aktywacji urządzeń dla poszczególnych wzorców technologicznych, których detekcja opisana jest w rozdziale 4., wprowadzone do modelu materiałowego jako fragment opisu produkowanych wariantów, powiązane są na poziomie tego modelu ze składowymi technologiami (poprzez wskazanie jakie urządzenia i z jakimi nastawami produkują poszczególne warianty produktów końcowych). Inne specyfiki procesów produkcyjnych mogą umożliwić wykorzystanie identyfikatora urządzenia jednak w przypadku stanowiska DC40 nie było takiej możliwości z uwagi na wykorzystanie prostych komponentów produkcyjnych, nieposiadających ani interfejsów cyfrowych, ani konieczności ich unikalnego identyfikowania na poziomie sterownika PLC.

Dzięki powiązaniu konkretnego urządzenia z rezultatami działania metody detekcji anomalii możliwe jest wsparcie wiedzą technologiczną przy próbie wyeliminowania tych problemów. Wskazanie na urządzenie wykonawcze zamodelowane zgodnie ze standardem ANSI/ISA-95 i w oparciu o język AML ukazuje powiązania pomiędzy składowymi tego urządzenia i odniesienia do zasobów zewnętrznych, jak również prezentuje lokalizację analizowanego urządzenia w strukturze instalacji. Przykład powiązania tych elementów z urządzeniem typu ssawka na stanowisku DC40 przedstawiony został na rysunku 3.3.2. Rysunek prezentuje powiązanie adresu sterującego pracą urządzenia Q20.6 z kartą wyjść sterujących PLC (niebieskie połączenia *Internal Link*), odniesienie do dokumentacji producenta urządzenia (adres URI) oraz wykorzystanie dwóch atrybutów opisujących minimalny i maksymalny czas aktywacji siłownika.

Możliwym rozwinięciem podejścia wspierającego detekcję i obsługę wykrytych błędów technologicznych jest zapewnienie możliwości dostępu do wartości procesowych zmieniających się w czasie. Mechanizm subskrypcji OPC UA Data Access zapewnia najbardziej aktualne wartości procesowe dla danego węzła. Dla analizowanego przypadku będzie to całkowity czas aktywacji urządzenia wykonawczego. Niestety mechanizm ten nie zapewnia informacji o historii zmian wartości procesowych, które mogą być istotne, np. dla potrzeb detekcji postępujących degradacji komponentów instalacji, np. wydłużającego się czasu pracy wskazującego na możliwe uszkodzenie siłownika pneumatycznego. Dla

tego celu konieczne jest użycie mechanizmu OPC Historical Access, które wymaga jednak zastosowania zewnętrznego repozytorium zarchiwizowanych danych produkcyjnych, ponieważ OPC UA jest jedynie interfejsem komunikacyjnym i nie posiada samodzielnych zdolności do przechowywania danych.



Rysunek 3.3.2. – powiązanie wpisu w modelu technologicznym z daną procesową przez adres sterujący

Dla tego celu możliwe jest użycie bazy danych, w której przechowywane są wartości procesowe istotne dla analiz. Ponieważ nie wszystkie wartości tego wymagają, interfejs OPC UA umożliwia wybranie odpowiedniego atrybutu historyzacji.

3.4. Podsumowanie

Wykorzystanie modeli informacyjnych instalacji i materiałów wykorzystywanych do produkcji w połączeniu z przestrzenią adresową standardu OPC UA umożliwia powiązanie wiedzy o technologii wykonania instalacji przemysłowej z metodami analizy efektywności i jakości pracy stanowisk wykonawczych. Ponieważ nowoczesne systemy realizacji produkcji wymagają obsługi produkcji krótkoseryjnej i masowo-dostosowanej, a więc wielowariantowej, dotychczasowe modele informacyjne, np. popularny model ANSI/ISA-95 nie są w stanie zapewnić takiego powiązania, skupiając się jedynie na strukturze instalacji lub zarządzaniu procesami operacyjnymi, a nie na integrowaniu wiedzy o jakości pracy urządzeń lub realizacji procesu z technologią wykonania instalacji. Zaproponowane przez autora podejście przez zamodelowanie fizycznej struktury instalacji w języku AutomationML i wykorzystanie właściwości hierarchii instancji, zastosowanie modelu informacyjnego materiałów do opisanego analizowanych zasobów technologicznych oraz zastosowanie interfejsu OPC UA do stworzenia platformy wymiany danych i powiązania za ich pomocą różnych form opisu sprzętu i procesu, doprowadziło do powiązania z rezultatami detekcji źródeł anomalii i umożliwiło integrację wiedzy o technologii z monitoringiem danych procesowych. Podejście to jest zgodne z architekturą RAMI4.0 – aktualne wartości procesowe oraz wartości historyczne opisane przez model struktury instalacji i model informacyjny materiałów (ang. Material Information) dostępne są z wykorzystaniem odpowiednich usług OPC UA, dzięki czemu budowany jest cyfrowy bliźniak instalacji.

Ograniczeniem zaproponowanego podejścia jest bazowanie na istniejących standardach i modelach, cechujących się ograniczeniami w specyficznych warunkach produkcyjnych. Architektura RAMI 4.0 bazuje na hierarchicznych modelach wyposażenia, personelu i zasobów, wywodzących się ze standardu ISA-95. Stosowanie ich, zwłaszcza w niewielkich i ograniczonych środowiskach produkcyjnych, znacząco komplikować może implementację systemów, głównie z uwagi na złożoność tych modeli. Dodatkowym ograniczeniem jest konieczność definiowania indywidualnego instancyjnego modelu infrastruktury stanowiska produkcyjnego, który w przypadku badań autora opracowany został w formacie AutomationML. Choć czynność ta może być usprawniona przez podejście obiektowe i stosowanie wspólnych instancji, to nadawanie wartości atrybutów i tworzenie powiązań pomiędzy szczegółowymi składowymi konieczne jest do indywidualnego przeprowadzania dla każdej opisywanej instalacji.

Prace badawcze skoncentrowane wokół zagadnień opisanych w tym rozdziale zostały przedstawione i opublikowane w czasopismach naukowych i podczas konferencji:

- Cupek, R., Ziebinski, A., Drewniak, M., & Fojcik, M. (2019). Knowledge integration via the fusion of the data models used in automotive production systems. *Enterprise Information Systems*, 13(7-8), 1094-1119.
- Cupek, R., Drewniak, M., Ziebinski, A., & Fojcik, M. (2019). “Digital Twins” for highly customized electronic devices—Case study on a rework operation. *IEEE Access*, 7, 164127-164143.
- Cupek, R., Drewniak, M., & Ziebinski, A. (2019, October). Information models for a new generation of manufacturing systems—a case study of automated guided vehicle. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 858-864). IEEE.
- Cupek, R., Ziebinski, A., & Drewniak, M. (2017, March). An OPC UA server as a gateway that shares CAN network data and engineering knowledge. In *2017 IEEE International Conference on Industrial Technology (ICIT)* (pp. 1424-1429). IEEE.
- Cupek, R., Ziębiński, A., Drewniak, M., & Fojcik, M. (2017, September). Application of OPC UA Protocol for the Internet of Vehicles. In *International Conference on Computational Collective Intelligence* (pp. 272-281). Springer, Cham.

4. METODA AUTOMATYCZNEGO WYKRYWANIA WZORCÓW TECHNOLOGICZNYCH Z WYKORZYSTANIEM KLASTERYZACJI METODĄ k -MEANS

Niniejszy rozdział opisuje badania nad rozwiązaniem dwóch problemów przedstawionych w rozdziale 1.: problemu automatycznego wykrywania profili produkcyjnych z wykorzystaniem typowo stosowanych wielkości procesowych i bez znajomości technologii oraz problemu selekcji i agregacji danych, na podstawie których prowadzone mogą być analizy.

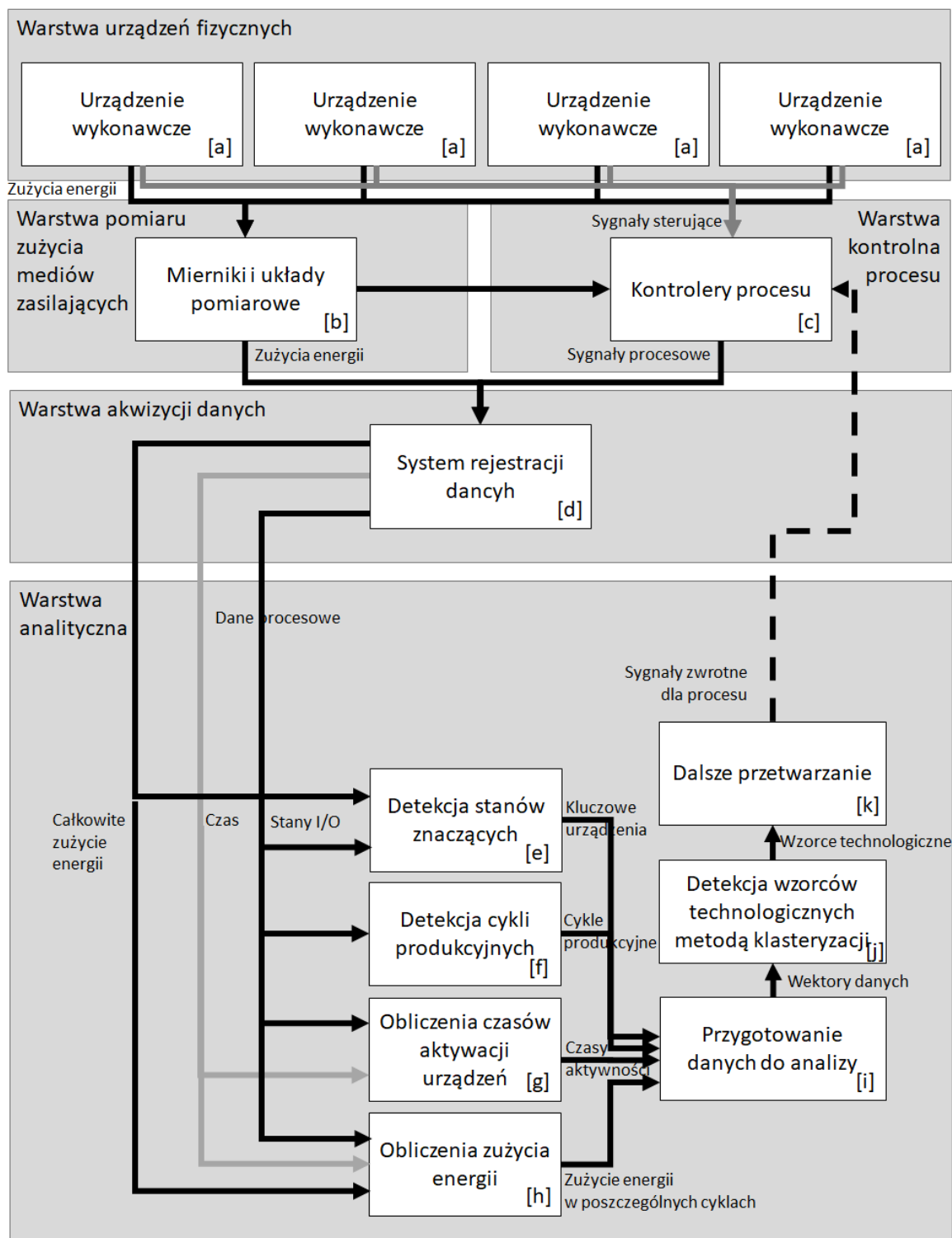
W rozdziale 3. przedstawione zostało podejście do powiązania informacji o technologii wykonania instalacji produkcyjnej z danymi procesowymi, które mogą zostać wykorzystane w analizach jakości i efektywności pracy tej instalacji. W szczególności możliwe jest powiązanie technologii z wielkościami procesowymi, definiującymi wzorce technologiczne. Ponieważ instalacje przemysłowe mogą być wyposażone w różne zestawy urządzeń i cechować się różną specyfiką pracy, poszczególne składowe wektora definiującego wzorzec technologiczny nie zawsze muszą występować, np. dla procesu montażowego, realizowanego przez ramię zrobotyzowane niekoniecznie musi wystąpić zużycie zasobów produkcyjnych. Z tego powodu, choć wzorzec może służyć jako uniwersalny sposób reprezentacji wariantu realizowanego w trakcie procesu, różne instalacje posiadać będą oddzielne wzorce. Jedną z możliwości zastosowania wzorców jest wykorzystanie ich jako referencyjnych wartości definiujących jakość pracy instalacji przy pojedynczym produkcie. Dzięki temu możliwe jest m.in. wykrywanie defektów jakościowych, przejawiających się zwiększonym zapotrzebowaniem na wybrane surowce produkcyjne podczas bieżącej pracy. Drugą możliwością jest wykorzystanie ich w budowaniu mechanizmów diagnostycznych stanowisk, np. poprzez wykrywanie zwiększonych zużyc energii lub wydłużonych czasów wysterowań pojedynczych urządzeń wykonawczych, a przez to wczesne wykrywanie nieprawidłowości i zapobieganie awariom. Dodatkowo, możliwe jest poszukiwanie miejsc do optymalizacji zużycia energii lub zasobów, np. poprzez redukcję czasu wysterowań wybranych urządzeń. Ograniczenia w stosowaniu wzorców produkcyjnych mogą wynikać z różnych powodów. Najczęstszym z nich jest brak rejestrowania odpowiednich danych procesowych, które umożliwiają ich wyznaczenie. Aby pozyskać informacje o statusach aktywności i czasach pracy urządzeń wykonawczych konieczne jest dostosowanie procedur sterujących kontrolerów procesu,

np. sterowników PLC tak, aby zapewnić rejestrację sygnałów. Niejednokrotnie oznacza to ingerencję w działający algorytm sterujący, co w wielu wypadkach jest kłopotliwe lub, przez wrażliwość procesu, niemożliwe. Równie częsty jest brak odpowiedniego doposażenia stanowisk produkcyjnych w układy pomiarowe, np. do rejestracji energii. Większość systemów produkcyjnych nie monitoruje ich zużycia, realizując po prostu zlecenia produkcyjne wynikające z kontraktów tak, aby spełniać oczekiwania produktowe i parametry narzucone przez klientów. Choć koszt zużycia energii niezbędnej do produkcji jest jednym z najważniejszych kosztów w procesie wytwarzania, zwykle nie jest on zdefiniowany ogólnie dla indywidualnych produktów, a szacowany jest dopiero na poziomie realizacji zadań produkcyjnych. Nawet jeżeli informacja o profilach produkcyjnych jest dostępna dla systemów monitorujących pracę instalacji, nie jest ona przydatna do analizy efektywności zużycia energii. Wynika to przede wszystkim ze specyfiki wykonania poszczególnych instalacji, np. długości przewodów zasilających, warunków pracy instalacji czy stopnia zużycia narzędzi i urządzeń wykonawczych. Skutkować to może wystąpieniem znaczących różnic energetycznych w produkcji jednakowego wariantu produktu, nawet w obrębie pojedynczego zakładu produkcyjnego (ten sam produkt wytwarzany na różnych instalacjach wymagać będzie różnego zapotrzebowania na energię). Wreszcie, ograniczeniem w stosowaniu wzorców mogą być plany produkcyjne, które układane są zwykle tak, aby zapewnić jak największe wykorzystanie dostępnych do produkcji zasobów (personelu, materiałów) i dążyć do jak najbardziej efektywnego wykorzystania stanowisk produkcyjnych. Skutkuje to częstym przemieszaniem wariantów produkowanych wyrobów, które znacząco mogą różnić się od siebie stopniem złożoności wykonania. Efektem tego przemieszania może być zaburzenie wartości cech definiujących wzorce, wynikające np. ze zmian nastaw, recalibracji i ponownego nasycania urządzeń (osiągania gotowości do pracy) i wywołujące ich nieprzewidywalność. Wszystkie te czynniki powodują, że wykrywanie wzorców produkcyjnych, między innymi dla potrzeb optymalizacji zużycia energii, nie może odbywać się przez odgórne pobieranie informacji o nich z nadrzędnych systemów. Jednocześnie powinno być ono możliwie najbardziej uniwersalne i realizowane niezależnie od stosowanych technologii tak, aby zapewnić aplikowalność dla różnego rodzaju zakładów i instalacji.

Odpowiedzi systemów kontrolno-pomiarowych mogą być wartościowym źródłem informacji i zapewniać możliwość poprawy procesu. Z tego powodu właściciele instalacji

coraz częściej dążą do monitorowania sygnałów sterujących i zużyć, a właśnie wzorce technologiczne z powodzeniem mogą zapewniać rolę nośnika dla tych informacji. Niestety, tak z punktu technologii, jak i technik informatycznych rozpoznawanie i klasyfikacja tych wzorców mogą być bardzo trudne, głównie ze względu na stopień złożoności instalacji lub mnogość oraz pojawianie się wciąż nowych wariantów produkcyjnych. W rozdziale 2.6. opisane zostały podejścia do wykrywania profili produkcyjnych i detekcji wariantów z wykorzystaniem różnych technik. Ponieważ zaprezentowane podejścia nie bazują na budowaniu informacji o wariantach w oparciu o wzorce technologiczne lub nie zakładają możliwości dynamicznego pojawiania się nowych wariantów produkcji, autor w swoich badaniach skupił się na mechanizmach przygotowania cech tych wzorców oraz nad metodami rozpoznawania, grupowania i automatycznej detekcji nowych ich typów. Wykrywanie wariantów oparte zostało o wzorce technologiczne zdefiniowane przez czasy pracy stacji, czasy pracy urządzeń wykonawczych i energię zużywaną w trakcie pracy. Wybór tych atrybutów podyktowany był specyfiką produkcji dyskretniej, w której niejednokrotnie nie występuje zużycie dodatkowych zasobów, a w której zawsze zużywany jest co najmniej jeden typ medium zasilającego proces, zwykle energia elektryczna lub sprężone powietrze. Pozostałe atrybuty, tj. czasy pracy stacji i urządzeń, są ogólnodostępnymi wartościami do wyznaczenia na podstawie sygnałów procesowych generowanych przez kontroler, co zapewniać ma odpowiednią uniwersalność i skalowalność rozwijanego rozwiązania.

Schemat ideowy metody do automatycznego wykrywania wzorców technologicznych w oparciu o dane procesowe i dane o zużyciu mediów zasilających stanowisko zaproponowanej przez autora, przedstawiony został na rysunku 4.1. Przedstawia on kroki wykonywane od pozyskiwania sygnałów procesowych i statusowych urządzeń wykonawczych po wykrywanie i identyfikację wzorców technologicznych wraz z definiującymi je wartościami oraz możliwości ich dalszego wykorzystania.



Rysunek 4.1. – schemat metody do automatycznego wykrywania wzorców technologicznych

Warstwa urządzeń fizycznych obejmuje urządzenia wykonawcze (oznaczone na schemacie symbolem [a]) wykorzystywane w trakcie produkcji, na podstawie których budowane mają być wzorce. Do warstwy należeć mogą wszelkiego rodzaju napędy, ramiona zrobotyzowane, osie numeryczne, wkrętarki czy siłowniki, które sterowane są

z poziomu warstwy kontrolnej procesu, do której należeć mogą np. sterowniki PLC i komputery klasy IPC.

Warstwa pomiaru zużycia mediów zasilających zawiera urządzenia pomiarowe, rejestrujące zużycie energii i mediów [b], które zasilają stanowisko w trakcie jego pracy. Mogą to być przepływomierze rejestrujące chwilowe zużycia powietrza, analizatory parametrów sieciowych mierzące zużycie energii elektrycznej czy przepływomierze wskazujące zużycie chwilowe chłodziw, olejów czy silikonu. Ponieważ układy pomiarowe mogą pracować niezależnie od warstwy kontrolnej procesu, rejestrowane dane mogą być zarówno przekazywane do układów sterujących, [c] jak i do zewnętrznych systemów rejestrujących dane, np. systemów SCADA czy narzędzi analitycznych.

Warstwa akwizycji danych odpowiedzialna jest za rejestrowanie i gromadzenie sygnałów procesowych oraz pomiarowych, dostarczanych z niższych warstw oraz przekazywanie tych danych do dalszego przetwarzania. Składa się ona z rozwiązania wyposażonego w zestaw narzędzi do komunikacji z kontrolerem procesu i urządzeniami pomiarowymi, repozytorium do przechowywania danych procesowych oraz zestaw metod do przekazywania tych danych do dalszego przetwarzania [d]. Może to być komputer klasy PC z odpowiednio przygotowanymi tabelami baz danych, komputer typu embedded przechowujący dane w postaci plików lub dowolne inne rozwiązanie wykorzystujące nierelacyjne bazy danych, np. Cassandra czy ELK. W przypadku badań prowadzonych przez autora, warstwa akwizycji danych przygotowana została w oparciu o zewnętrzny komputer PC zlokalizowany przy stanowisku produkcyjnym, komunikujący się ze sterownikiem PLC za pomocą protokołu TCP/IP oraz zbierający dane z wykorzystaniem plików .csv oraz serwera OPC UA i udostępniający te dane do dalszego, lokalnego przetwarzania. Do danych tych należą: stempel czasowy rejestrowanych pomiarów generowany przez sterownik PLC, wartość chwilowa lub całkowita przyrostowa zużycia mediów zasilających, rejestrowana przez urządzenia pomiarowe dostarczające wskazania do sterownika PLC oraz stany wejść i wyjść cyfrowych sterownika PLC, odpowiadające aktywnościom poszczególnych urządzeń wykonawczych. W celu zapewnienia prawidłowości analiz i nauki modelu wykrywania wzorców niezbędne jest, aby gromadzone dane zawierały prawidłowe, niezaburzone anomaliami cykle produkcyjne, w trakcie których wykorzystywane są określone zestawy urządzeń wykonawczych, pracujących zgodnie z zaprogramowaną dla wariantu produkcyjnego sekwencją lub kombinacją. Cykle te odpowiadają pracy na wariantach produkcyjnych, które mają być

identyfikowane. Jeżeli analizie poddane zostaną cykle zaburzone, metoda wykryje nadmiarową, a więc nieprawidłową liczbę wzorców technologicznych.

Warstwą odpowiedzialną za przetwarzanie danych procesowych jest warstwa analityczna. Ze względu na specyfikę uruchamianych metod, wymaga ona odpowiednich zasobów sprzętowych i mocy obliczeniowej, zdolnych do przechowywania odpowiednio dużych zbiorów danych i przeprowadzenia operacji analitycznych. Z tego powodu nie może to być klasyczny sterownik PLC czy kontroler oparty o architekturę czasu rzeczywistego, a zewnętrzny system wykorzystujący komputer klasy PC lub klaster obliczeniowy, z uruchomionym środowiskiem obliczeniowym, np. rozwiązaniem przygotowanym w otwartym środowisku programistycznym, rozproszonym systemem analitycznym, np. Apache Spark, narzędziach matematycznych Matlab lub RapidMiner czy frameworkach MOA lub WEKA. Podobnie jak warstwa akwizycji danych, również warstwa analityczna zbudowana przez autora dla potrzeb badań oparta została na lokalnym komputerze PC (tej samej jednostce) z uruchomionym środowiskiem Matlab. Zgromadzone dane przetwarzane są w czterech krokach przygotowawczych: w wyznaczaniu stanów znaczących wchodzących w skład cykli produkcyjnych [e], wykrywaniu momentów rozpoczęcia i zakończenia cykli produkcyjnych [f], wyliczeniu zagregowanej wartości czasu aktywności poszczególnych urządzeń wykonawczych w pojedynczych cyklach produkcyjnych [g] oraz wyliczeniu zagregowanej wartości energii pochłanianych w trakcie tych cykli [h]. Pierwszy z kroków [e], to wyznaczanie stanów znaczących w analizowanych cyklach produkcyjnych. Ma on na celu określenie istotnych urządzeń wykonawczych, których wykorzystanie w różnych wariantach produkcyjnych może wpływać w znaczący sposób na czasy i zużycia energii w wykrywanych wzorcach technologicznych. Etap ten stanowi dodatkowo element filtracji danych poprzez wyłonienie tych cech, które będą definiować strukturę wykrywanych w dalszym etapie wzorców. Drugi krok [f], stanowi wykrywanie cykli produkcyjnych. Jest on konieczny z dwóch powodów: po pierwsze, stanowi niezbędne uzupełnienie do określenia zestawu urządzeń wykonawczych podejmujących pracę w ramach realizacji danego wariantu produkcyjnego oraz sekwencji, w jakiej są aktywowane, a po drugie w celu zdeterminowania odpowiedniego okna, dla którego wyliczane będą zagregowane wartości zużycia energii i czasów aktywności urządzeń, wyznaczane w pozostałych krokach [g][h]. Wiedza o sekwencji pozwoli na rozróżnienie poszczególnych wzorców produkcyjnych na podstawie aktywności tych urządzeń, ale ponieważ może być ona

niewystarczająca, zwłaszcza w przypadkach produkcji za pomocą stałego zestawu urządzeń, jednak charakteryzującego się różnymi parametrami ich pracy, konieczne jest jej uzupełnienie o czasy aktywności oraz pochłaniane energie. Obliczenia tych wartości realizowane są w krokach: trzecim [g] oraz czwartym [h].

Drugim etapem jest wyłonienie zestawu danych, które opiszą wykryte cykle produkcyjne, i zostaną dodatkowo uzupełnione o wyliczone we wcześniejszym etapie wartości zagregowane [i]. Dane zebrane w postaci wektorów odpowiadających pojedynczym cyklom produkcyjnym przekazywane są do trzeciego etapu [j].

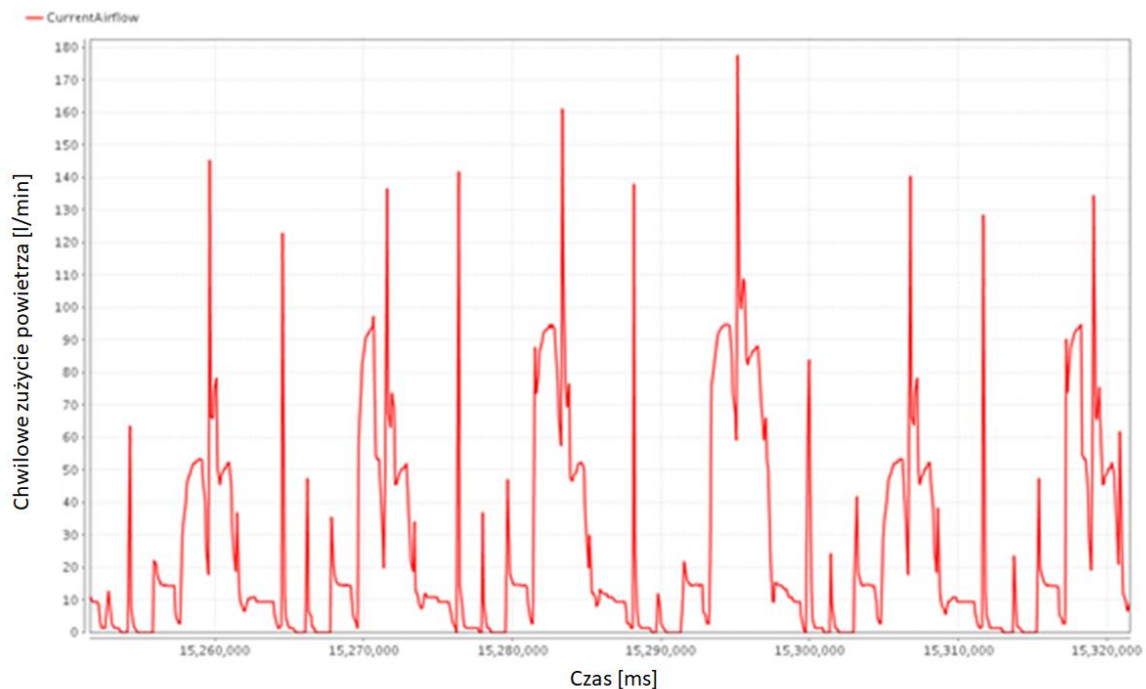
Trzeci etap [j] to wykrywanie wzorców technologicznych poprzez klasteryzację danych wielowymiarowych z wykorzystaniem algorytmu *k*-Means, zasilanego wektorami opisującymi zestawy wykorzystywanych urządzeń wykonawczych, czasy ich aktywności i energię pochłanianą w trakcie cykli produkcyjnych.

Zbudowane informacje o wzorcach mogą posłużyć w etapie dalszego przetwarzania [k], m.in. jako referencyjne wartości w technikach wykrywania anomalii w pracy stanowisk, które opisane są w dalszych rozdziałach pracy.

4.1. Detekcja cykli produkcyjnych i stanów znaczących wykorzystywanych w wykrywaniu wzorców technologicznych

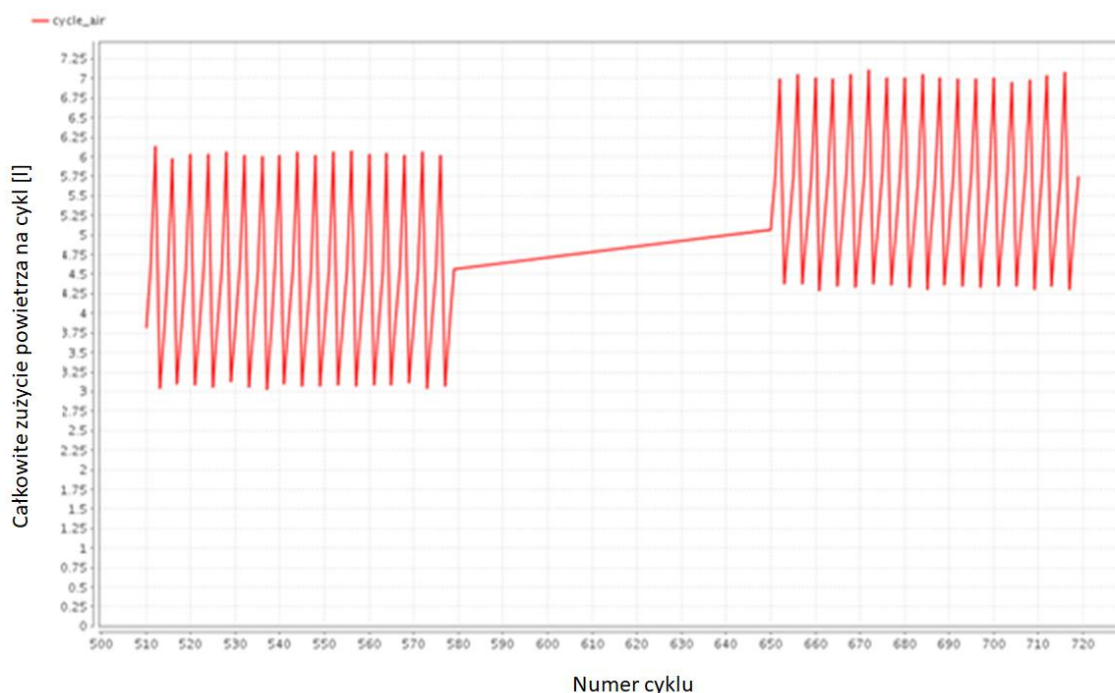
Pierwszy, prowadzony przez autora, zestaw badań dotyczył przygotowania danych procesowych do dalszej analizy i obejmował rozwój kroków [e-h] przedstawionych na rysunku 4.1. Zakres zrealizowanych prac doprowadził do wyznaczenia istotnych cech obserwacji oraz odpowiedniego zagregowania ich wartości, tak, aby umożliwić dalsze wykorzystanie w rozpoznawaniu wzorców technologicznych. Badania nad metodami prowadzone były z wykorzystaniem eksperymentalnego stanowiska produkcyjnego DC40, opisanego w załączniku 1.

Dane o zużyciu medium zasilającego dostarczane były w postaci wskazania chwilowego przepływu powietrza w interwałach co 100 ms. Przykładowy fragment zarejestrowanego chwilowego zużycia powietrza w czasie przedstawiony jest na rysunku 4.1.1.



Rysunek 4.1.1. – fragment zarejestrowanego zużycia chwilowego podczas pracy stanowiska DC40

Zaprezentowana próbka danych obejmuje sześć cykli produkcyjnych, których powtarzalność jest możliwa do zaobserwowania przez wizualną ocenę powtarzalności kształtów przebiegów zużycia. Typowy czas pojedynczego cyklu produkcyjnego wynosił pomiędzy 10 a 13 sekund. Nie można jednak określić tego, że zaprezentowane sześć cykli obejmowało produkcję czterech różnych wariantów produkcyjnych. W celu prezentacji tych wariantów, wstępnej ocenie można poddać całkowite zużycie sprężonego powietrza w trakcie realizacji poszczególnych cykli produkcyjnych. Wielkość ta przedstawiona została na rysunku 4.1.2.



Rysunek 4.1.2. – całkowite zużycie powietrza w zarejestrowanych cyklach produkcyjnych nr 520-720 podczas pracy stanowiska DC40

Wskazania zużycia całkowitego, pokazują, że spośród około 130 zarejestrowanych cykli produkcyjnych, pierwszy wariant zużywał około 3 litry powietrza, drugi 6, trzeci 4,5, a czwarty 7. Liniowy przyrost energii całkowitej pomiędzy dwoma zbiorami wariantów odpowiada czasowi przezbrojenia stanowiska pomiędzy produkcją w 1. i 2. wariantcie a tą w 3. i 4, w trakcie którego prowadzone były testy stanowiska i realizowane kolejne cykle, jednak podczas którego nie rejestrowano danych procesowych (odpowiada to skokowi energii pomiędzy ostatnim zarejestrowanym cyklem z pierwszej grupy i pierwszym cyklem z drugiej). Informacja ta nie jest możliwa wprost do wyciągnięcia na podstawie wartości chwilowej zużycia, stąd też do dalszych analiz autor podjął założenie, że metoda wykorzystywać będzie wartości całkowite zużyć w wybranych okresach pracy stacji, w tym przypadku: na poszczególny cykl produkcyjny. Do ich określenia niezbędne jest wyznaczenie różnicy pomiędzy kumulacyjną, przyrostową wartością zużycia na końcu i na początku analizowanych cykli.

Krokiem pośrednim do rozwiązania problemu detekcji wzorców technologicznych może być analiza cykli ze względu na sekwencje aktywacji urządzeń wykonawczych. Stanowisko DC40 składa się z dziewięciu głównych odbiorników pneumatycznych, więc kodowanie stanów pracy tych urządzeń możliwe jest na dwóch bajtach. Bajty wyjściowe kontrolowane przez sterownik PLC, do których podłączone były pneumatyczne urządzenia

wykonawcze posiadały adresy BQ20 i BQ21, stąd też do zestawu danych wejściowych konieczne było rozpoczęcie ich obserwacji. Przykład sekwencji składającej się ze stanów zdefiniowanych przez jednocześnie wysterowane urządzenia wykonawcze przedstawiony został w tabeli 4.1.1. Prezentuje ona zarejestrowane cykle w przedziale od 650 do 675, w których kodowanie poszczególnych stanów i sekwencji, w jakich występują, zestawiono w kolumnie oznaczonej „Sekwencja”.

Tabela 4.1.1. – wybrane cykle produkcyjne stanowiska DC40, zestawiające aktywne stany wyjść cyfrowych PLC

Numer cyklu	Zużycie powietrza [l]	Sekwencja	Czas PLC [ms]
650	5.062	37-0;37-1;103-0;102-0;38-0;42-0;154-0;26-0;42-0;0-38-0;	15138946
651	5.754	37-0;102-1;102-1;38-0;42-0;154-0;26-0;42-0;38-0;	15150264
652	6.987	37-0;101-1;102-1;38-1;42-1;154-0;26-0;42-0;38-0;	15162520
653	4.371	37-0;103-0;101-0;38-0;42-0;154-0;26-0;42-0;38-0;	15124492
654	5.063	37-0;37-1;10.1.-0;102-0;38-0;42-0;1154-0;26-0;38-0;	15186192
655	5.733	37-0;101-1;102-1;38-0;42-0;154-0;26-0;42-0;38-0;	15196120
656	7.047	37-0;101-1;102-1;38-1;42-1;42-1;154-0;26-0;42-0;38-0;	15209924
657	4.375	37-0;103-0;102-0;38-0;42-0;154-0;26-0;42-0;38-0;	15252559
658	5.072	37-0;37-1;102-0;38-0;42-0;170-0;154-0;26-0;42-0;42-0;38;	15234123
659	5.744	37-0;101-1;102-1;38-0;42-0;170-0;154-0;26-0;42-0;38-0;	15245929
660	7.007	37-0;101-1;102-1;38-1;42-1;42-1;154-1-0;26-0;42-0;38-0;	15257559
661	4.287	37-0;102-0;38-0;42-0;170-0;26-0;42-0;38-0;	15269421
662	5.076	37-0;102-1;102-0;38-0;42-0;154-0;26-0;42-0;38-0;	15281292
663	5.766	37-0;102-1;102-1;38-0;42-0;154-0;26-0;42-0;38-0;	15293109
664	6.982	37-0;102-1;38-1;42-1;42-1;154-0;26-0;42-0;38-0;	15304695
665	4.350	37-0;102-0;38-0;42-0;170-0;154-0;26-0;42-0;38-0;	15312042
666	5.062	38-0;37-1;101-0;102-0;38-0;42-0;154-0;26-0;42-0;38-0;	15330932
667	5.262	37-0;102-1;102-1;38-0;37-0;154-0;26-0;42-0;38-0;	15342646
668	2.044	37-0;101-1;102-1;38-1;46-1;42-1;42-1;154-0;26-0;42-0;38-0;	15254354
669	4.335	37-0;102-0;38-0;42-0;170-0;154-0;26-0;42-0;38-0;	15366452
670	5.040	37-0;37-1;103-0;102-0;38-0;42-0;154-0;26-0;58-0;42-0;38-0;	15390000
671	5.258	37-0;103-1;102-1;102-1;38-0;42-0;170-0;154-0;26-0;42-0;38-0;	15390000
672	7.104	37-0;101-1;102-1;38-1;42-1;42-1;154-0;26-0;42-0;38-0;	15401209
673	4.381	37-0;102-0;38-0;42-0;170-0;154-0;26-0;42-0;38-0;	15413655
674	5.099	37-0;37-1;102-0;38-0;42-0;170-0;154-0;26-0;58-0;42-0;38-0;	15425366
675	5.222	37-0;101-1;102-1;38-0;42-0;154-0;26-0;42-0;38-0;	15432025

Poszczególne stany oddzielone są znakami przecinka, podczas gdy poszczególne bajty wyjściowe znakami myślnika. W zaprezentowanym zestawie danych można zauważyć, że co czwarty zarejestrowany cykl cechuje się podobnym zużyciem powietrza, co dodatkowo potwierdzone zostało na etapie analizy wstępnej przez założenia

technologiczne dla stanowiska, realizującego cztery, powtarzające się w zapętłony sposób warianty produkcyjne. Co jednak jest dodatkowo zauważalne to fakt, że pomimo czterech powtarzalnych ze względu na zużycie energii wariantów produkcyjnych, w kolumnie sekwencji występuje aż 16 powtarzających się zestawów stanów. Powodem występowania większej liczby zestawu stanów niż faktycznych profili produkcyjnych, które powinny im odpowiadać, jest sposób rejestracji danych i realizacji algorytmu sterującego procesem. Sterownik PLC stanowiska posiada program sterujący, zbudowany w oparciu o układ kombinacyjny, umożliwiający jednoczesną aktywację zestawu wielu urządzeń na podstawie różnych warunków wejściowych. Powoduje to możliwość wystąpienia niewielkich różnic w sposobie wykonywania zadań, np. poprzez nieznaczne wydłużenie czasu aktywacji danego urządzenia w zależności od widoczności detalu produkcyjnego przez zestaw czujników. Dodatkowo, sterownik wystawia dane procesowe do rejestracji co 100 milisekund, co dodatkowo powoduje, że wysyłany jest zestaw danych „zatraśnięty” w stałych interwałach, które nie są w żaden sposób zsynchronizowane z cyklem programowym i produkcyjnym. Właśnie ten czynnik powoduje, że w zestawie stanów opisujących rejestrowane sekwencje występować mogą bardzo krótkie stany nieznaczające, będące w zasadzie stanami przejściowymi, które powinny być traktowane łącznie z wybranymi stanami sąsiedzkimi. Zjawisko to istotnie zwiększa trudność w analizie sekwencji i wymaga podjęcia dodatkowych kroków, które będą mieć na celu doprowadzenie do poprawności wyłaniania cykli i wzorców technologicznych.

Ponieważ nie wszystkie z opisanych powyżej stanów wnoszą istotne informacje o sposobie realizacji wariantów produkcyjnych przez stanowisko, konieczne jest wyłonienie tych, które faktycznie definiują sekwencje, w jakich realizowane były poszczególne cykle produkcyjne. W tym celu wykorzystano algorytm bazujący na powiązaniach pomiędzy wystąpieniami istotnych różnic w całkowitych zużyciach energii w poszczególnych stanach. Algorytm opisany został w [126] i zakłada wykorzystanie właściwości:

$$\sum D_S(X^k) > T_C \quad (1)$$

gdzie:

D_S = kumulacyjne zużycie energii

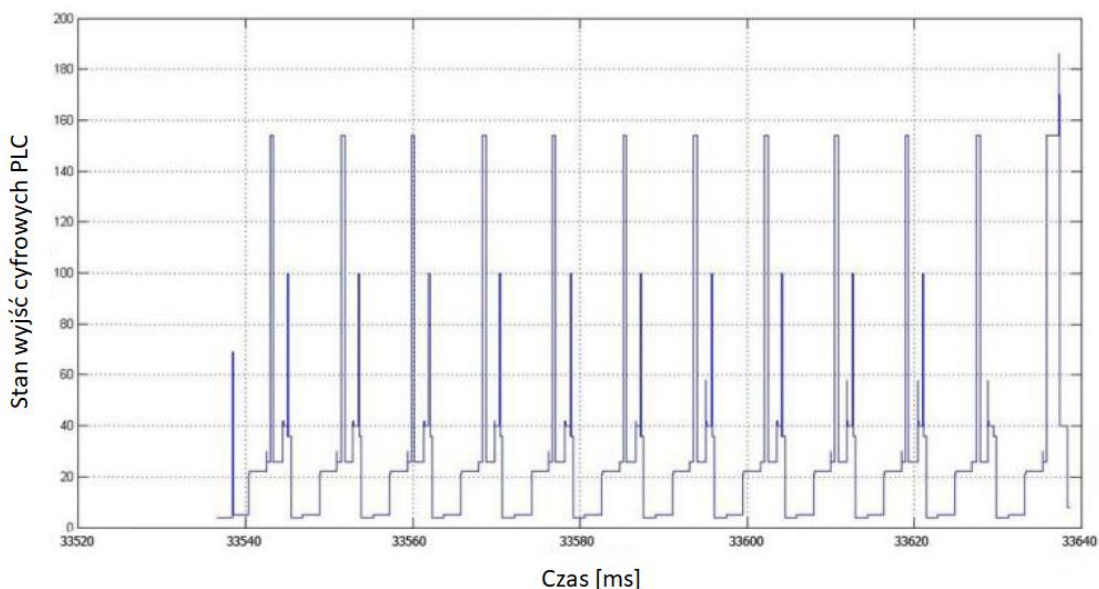
X^k = wektor zmian stanów wyjść cyfrowych sterownika PLC

k = stan wyjść cyfrowych sterownika PLC w zapisie dziesiętnym

T_C = wartość progowa wyznaczana z wzoru:

$$T_C = 10\% * \text{mediana}(\{\sum D_S(X^k)\}) \quad (2)$$

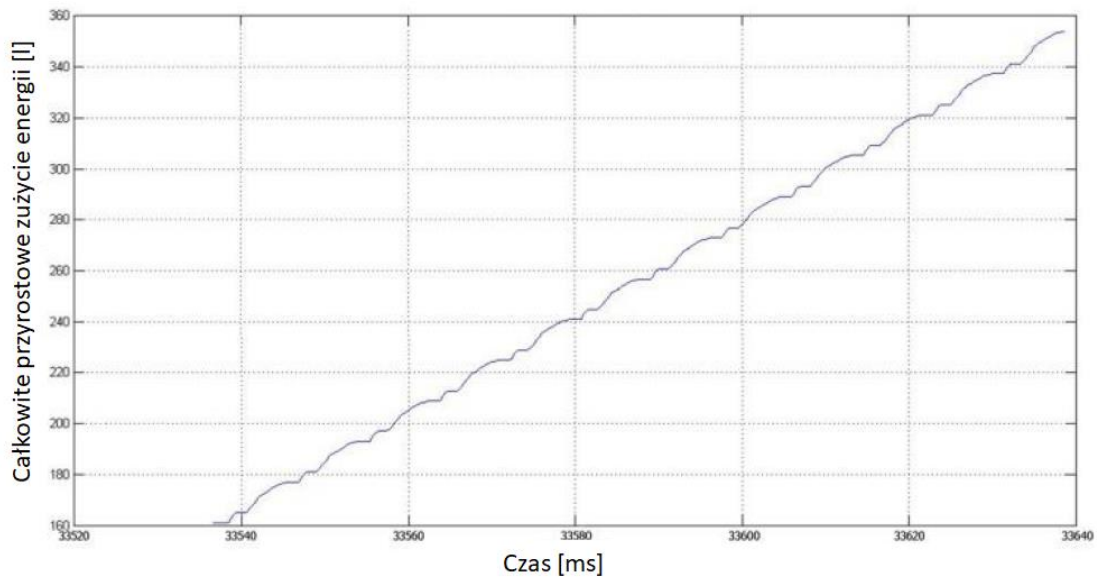
Wstępne dane zarejestrowane podczas pracy stanowiska DC40 obejmowały stemple czasowe, całkowitą kumulacyjną wartość zużywanego medium zasilającego oraz stany wyjść sterujących urządzeniami pneumatycznymi, które zebrane w dwubajt (słowo bitowe) reprezentowane były w postaci dziesiętnej. Przykład przebiegu zmian tych stanów przedstawiony został na rysunku 4.1.3., natomiast wykres przyrostowego zużycia powietrza przez stację w czasie jej pracy zaprezentowany został na rysunku 4.1.4.



Rysunek 4.1.3. – przebieg zmian wartości wyjść cyfrowych sterownika PLC na stanowisku DC40 w czasie

Przebieg zmian wartości wyjść cyfrowych aktywujących urządzenia pneumatyczne wskazuje na ich powtarzalność oraz na relacje, jakie pomiędzy nimi występują (np. stan oznaczony wartością 5 poprzedza zwykle stan oznaczony wartością 22). Podobnie jak dla wykresu z rysunku 4.1.1., przedstawiającego chwilowe wskazania przepływomierza, również tutaj widoczne są powtarzalne cykle produkcyjne. Jednocześnie zaobserwować można pewne nieregularności w długości trwania poszczególnych stanów (np. stanu 156 w ostatnim cyklu na wykresie) lub wystąpienie innych, krótkich i nadmiarowych (np. stanu 188 w ostatnim cyklu lub stanów 31 w cyklach numer 1-3). Z kolei wykres kumulacyjnego zużycia powietrza przedstawiony na rysunku 4.1.4. wskazuje na cykliczne, gwałtowniejsze

przyrosty zużycia energii, wskazujące na potencjalnie zwiększone zapotrzebowanie na sprężone powietrze w toku realizowanych cykli produkcyjnych.



Rysunek 4.1.4. – całkowite, przyrostowe zużycie powietrza podczas pracy stanowiska DC40

Dzięki zastosowaniu zależności (1) dla pierwszego z eksperymentalnych zbiorów danych zebranych podczas pracy stanowiska, wyznaczono referencyjny cykl pracy stanowiska, składający się z następujących stanów, oznaczonych jako znaczące:

$$PC = \{4, 5, 22, 26, 156, 40\}$$

W celu weryfikacji poprawności wyznaczonego cyklu pracy, zdekodowano stany zgodnie z tabelą 2 z załącznika opisującego stanowisko DC40, oraz określono wzajemną aktywność urządzeń wykonawczych w sekwencji. Aktywne urządzenia w kolejnych stanach opisane zostały w tabeli 4.1.2.

Tabela 4.1.2. – kombinacja aktywnych urządzeń wykonawczych w zależności od stanu

STAN [dec]	STAN [bin]	Aktywne urządzenia wykonawcze			
4	00000100	Wypychacz w przód	-	-	-
5	00000101	Wypychacz w przód	Podnośnik w górę	-	-
22	00010110	Wypychacz w przód	Podnośnik w dół	Zassanie	-
26	00011010	Wypychacz w tył	Podnośnik w dół	Zassanie	-
154	10011010	Wypychacz w tył	Podnośnik w dół	Zassanie	Stop w dół
40	00101000	Wypychacz w tył	Wydmuch	-	-

Jednocześnie z wyznaczaniem stanów znaczących niezbędne jest określanie czasów ich trwania oraz energii zasilających, które są w nich zużywane. Służyć to może zarówno analizie indywidualnych stanów, np. w celu porównania energii zużywanych w kolejnych cyklach, jak również do określenia zagregowanych wartości związanych z pojedynczymi cyklami pracy. Istotne są tu nie tylko powtarzające się sekwencje stanów, wskazujące na potencjalne początki cykli, ale również znaczniki czasowe umożliwiające wyznaczenie czasu trwania cykli wraz z pochłanianymi energiami. W tym celu, każdorazowo podczas określania stanu znaczącego, przetwarzana była informacja o czasie początku i końca tego stanu oraz energia kumulacyjna, rejestrowana na początku i na końcu wystąpienia. Obliczenie różnic pozwoliło na określenie zagregowanych wartości dla stanów i cykli, a następnie zostało wykorzystane w dalszych badaniach.

Otrzymane rezultaty są zgodne z oczekiwaniami: wykryta sekwencja odpowiada rzeczywistej kolejności, w jakiej aktywowane są urządzenia, zgodnie z narzuconą technologią produkcji w danym wariantcie oraz zgodnie z przygotowanym programem sterującym pracą stacji. Dla rozpatrywanego przypadku z pierwszego zbioru eksperymentalnego, urządzenie opisane jako „Wypychacz” realizuje jednorazowy ruch do przodu, a następnie pozostaje wycofane przy jednoczesnym podtrzymaniu wysterowania tak, aby zapobiec niekontrolowanemu ruchowi (budowa fizyczna stanowiska powodować może grawitacyjne opadanie tłoka tego siłownika, konieczne jest więc zapobieganie temu rozpozycjonowaniu). Urządzenie opisane jako „Podnośnik” realizuje ruch w górę po uprzednim wysunięciu „Wypychacza” oraz ruch w dół po aktywacji „Zassania”.

„Wydmuch” następuje dopiero po zakończeniu „Zassania”. Dodatkowo „Wypychacz” i „Podnośnik” są urządzeniami cechującymi się największymi zużyciami powietrza (głównie ze względu na duże komory powietrzne siłowników), a zapotrzebowanie to widoczne jest w postaci opisanych wcześniej gwałtowniejszych przyrostów zużycia kumulacyjnego, przedstawionego na rysunku 4.1.4.

O ile sekwencja stanów została wyznaczona i stwierdzono jej powtarzalność, jedno z urządzeń wykonawczych zostało pominięte: siłownik oznaczony jako „Prestop”. Ponieważ detekcja stanów znaczących wykorzystuje mechanizm przynależności do przedziału w zależności od wartości zużycia energii w trakcie trwania stanu, a urządzenie „Prestop” charakteryzuje się bardzo krótkim czasem wysterowania i bardzo niewielką pojemnością komory siłownika (niecałe 2 cm³), dobrana wartość 10% z mediany opisana we wzorze (2) odrzuciła stany, w których następuje wysterowanie tego urządzenia. Zwiększenie czułości metody prowadzić może do włączenia stanów, w których urządzenie „Prestop” pozostaje aktywne, jednak z punktu widzenia technologii, a w szczególności czasu wysterowania i zużywanej przez urządzenie energii, stany te są pomijalne.

Przedstawione metody detekcji cykli produkcyjnych i stanów znaczących umożliwiają zniwelowanie negatywnego zjawiska polegającego na wyodrębnieniu różnych cykli produkcyjnych, opisanych przez różne urządzenia wykonawcze aktywowane jednocześnie podczas sterowania kombinacyjnego. Dzięki nim możliwe jest znaczące uproszczenie danych surowych, które wprowadziłyby nadmierną złożoność wariantów obserwacji wykorzystywanych w dalszych etapach automatycznego wykrywania wzorców technologicznych. Mechanizmy te zapewniają dodatkowo istotny element przetwarzania danych: wstępnie agregują wartości, umożliwiając znaczące uproszczenie dalszych analiz, które wykorzystywać będą nie tylko dane wnoszące realną wartość do rozpatrywanych zagadnień (filtrując dane nieistotne ze względu na zużycie), ale także przyspieszające algorytmy uruchamiane w ramach etapu detekcji wzorców.

4.2. Detekcja wzorców technologicznych z wykorzystaniem algorytmu klasteryzacji *k*-Means

Kolejny zestaw prac związany był z wykrywaniem wzorców technologicznych, opartych o technikę klasteryzacji metodą *k*-Means. Jednocześnie z rozwojem metody prowadzone były badania nad przygotowaniem danych do klasteryzacji. Prace te obejmowały rozwój metod zaprezentowanych na rysunku 4.1. w postaci bloków [j-k] i stanowiły kontynuację prac opisanych w rozdziale 4.1.

Rozwiązanie automatycznego wykrywania wzorców technologicznych rozpoczęto od eliminacji wpływu negatywnego zjawiska nadmiarowych i zmultiplikowanych stanów występujących w sekwencjach, które mogą opisywać warianty produkcyjne. W tym celu zaproponowano model reprezentacji danych, który stanowi wektor zasilający etap detekcji wzorców technologicznych, oznaczony na rysunku 4.1. jako [j], a który zbudowany jest w oparciu o wektory cyklu stanowiska V_{MCn} (ang. Machine Cycle number), zdefiniowany jako:

$$V_{MCn} = \{V_{Cn}, Q1t, Q2t, \dots, Qmt\}$$

gdzie:

V_{MCn} = wektor pojedynczego cyklu produkcyjnego

n = numer cyklu

V_{Cn} = całkowite zużycie energii w cyklu produkcyjnym

$Q1t - Qmt$ = całkowity czas aktywności urządzenia wykonawczego $Q1 \dots Qm$

W celu zachowania uniwersalności rozwiązania i możliwości jego aplikacji dla różnych wariantów produkcyjnych wytwarzanych przez pojedynczą instalację, rozmiar wektora powinien być stały, a jego wielkość zależna od liczby urządzeń wykonawczych poddawanych analizie. W przypadku wystąpienia cyklu, w którym wybrane urządzenie pozostaje nieaktywne, odpowiadająca mu cecha wektora powinna pozostać wartością 0. Jednocześnie niezbędne jest traktowanie cech odpowiadających czasom aktywności urządzeń jako wzajemnie niezależnych od siebie – zależęć one mogą jedynie od technologii, która, zgodnie z założeniami badawczymi, pozostaje nieznaną. Zależność pomiędzy całkowitym zużyciem energii opisanym jako pierwsza cecha wektora, a czasami aktywności urządzeń stanowiącymi pozostałe cechy, może zostać opisana w postaci nieznanego współczynnika korelacji takiego, że:

$$V_{cn} = \sum_{i=1}^m f(Qit) \quad (3)$$

Funkcja $f(Qt)$ zależy od właściwości fizycznych urządzenia zużywającego energię. Urządzenia pobierać mogą energię w trybie ciągłym w trakcie pracy (np. jednostajnie obciążane napędy elektryczne po zakończeniu ramp akceleracji, dmuchawy po otwarciu zaworów), poprzez pobór skończonej ilości energii, nawet po zakończeniu ruchu (np. siłowniki z komorami powietrza) lub poprzez połączenie tych dwóch typów profili zużyć. Jeżeli Qt lub wartość funkcji $f(Qt)$ różnią się dla poszczególnych cykli produkcyjnych, pozwala to na stwierdzenie, że:

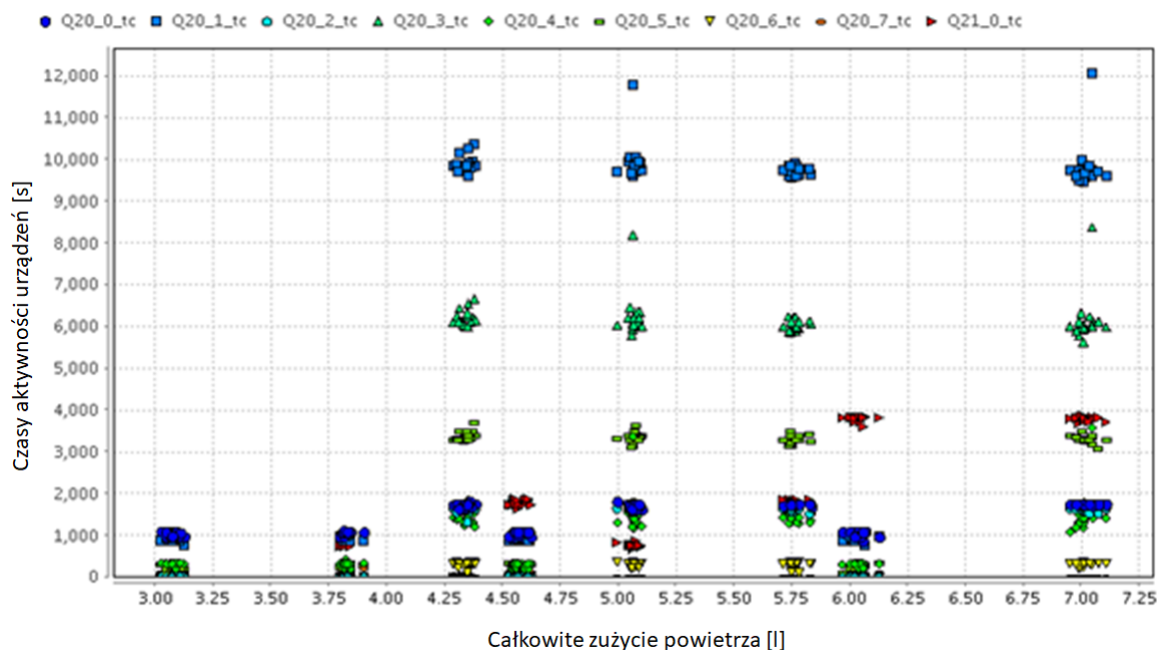
- Badane cykle różnią się ze względu na wzorzec produkcyjny lub,
- Zmieniły się parametry pracy urządzenia, np. na skutek zmiany nastaw lub na skutek wystąpienia anomalii w pracy układu lub,
- Zmianie uległy warunki procesowe, np. na skutek popełnienia błędu operatorskiego lub podczas awaryjnego zatrzymania pracy układu.

Model matematyczny funkcji $f(Qt)$ może być bardzo skomplikowany, a w praktyce niemożliwy do rozwiązania, głównie ze względu na różne kształty wspomnianych profili zużyć, w tym: różne poziomy granicy ustalenia sygnału, stromości zboczy narastania i opadania wartości sygnałów, czy różne czasy trwania stanu ustalonego. Z tego powodu autor zaproponował, aby nauka wartości poszczególnych cech wektorów opisujących cykle produkcyjne realizowana była w postaci obserwacji wartości referencyjnych zbieranych w trakcie niezaburzonych czasowo i energetycznie profili produkcyjnych. Obserwacje zebrane w postaci macierzy:

$$E_{MC} = \{V_{MC1}, V_{MC2}, \dots, V_{MCn}\}$$

mogą zostać następnie wykorzystane do analizy pracy układu, m.in. do wykrycia wzorców technologicznych, nauki nowych wzorców oraz detekcji anomalii.

Przykładowy zestaw danych zebranych ze stanowiska DC40, opisujący cykle produkcyjne przekształcone do postaci macierzy E_{MC} zaprezentowany został na rysunku 4.2.1.



Rysunek 4.2.1. – wizualizacja macierzy E_{MC} dla eksperymentalnego zbioru danych

Przedstawia on czasy aktywacji dziewięciu urządzeń wykonawczych, reprezentowanych przez dziewięć symboli graficznych o różnych kształtach i kolorach w funkcji całkowitego zużycia energii, w tym wypadku: powietrza w litrach. W czasie trwania eksperymentu, stanowisko produkowało w ośmiu różnych wariantach produkcyjnych przygotowanych tak, aby średnie zużycia poszczególnych urządzeń wykonawczych różniły się od siebie zależnie od wykorzystywanego profilu produkcyjnego. Zależność ta widoczna jest na rysunku, na którym możliwe jest wizualne rozróżnienie ośmiu wariantów, widocznych w postaci pionowych zgrupowań wartości, różniących się od siebie poziomem zużycia powietrza. Przypadek ten jest jednak dużym uproszczeniem, ponieważ zużycie na porównywalnym poziomie może występować jednocześnie w wielu, nawet znacząco różniących się od siebie, profilach produkcyjnych. Już na zaprezentowanym zbiorze widać, że profile numer 6 i 7, pomimo nieznaczącej różnicy w zużyciach energii (średnio 5.75 litra dla wariantu 6 oraz średnio 6 litrów dla wariantu 7) istotnie różnią się, nie tylko ze względu na czasy aktywacji urządzeń wykonawczych (np. urządzenie reprezentowane przez czerwone trójkąty pozostaje aktywne przez średnio 1.5 sekundy dla wariantu numer 6 oraz przez około 4 sekundy dla wariantu numer 7), ale nawet ze względu na zestaw wykorzystywanych w trakcie pracy urządzeń (wariant numer 6 wykorzystuje 8 urządzeń, podczas gdy wariant numer 7 jedynie 5 urządzeń). Nasuwa to wniosek, że choć możliwe jest, aby kombinacje czasów aktywacji urządzeń prowadziły do porównywalnego zużycia energii pomiędzy różnymi profilami

produkcyjnymi, to ciągle różnić się one mogą ze względu na średnie czasy wysteroowań urządzeń. Właśnie ta właściwość wykorzystana została w dalszych badaniach autora nad automatyczną detekcją profili produkcyjnych i wykrywaniem anomalii w pracy układów, opisanych w rozdziałach 4.3. oraz 5.

Podjęciem do rozwiązania problemu detekcji wzorców technologicznych jest analiza skupień poszczególnych cech wektorów V_{MCn} wewnątrz macierzy E_{MC} i wyłonienie grup, które charakteryzują się odpowiednim poziomem koncentracji wokół wartości referencyjnych tych cech.

Metodą zastosowaną do zgrupowania cykli w zbiory jest klasteryzacja k -Means, opisana w rozdziale 2.4.1. Ponieważ jej podstawowe warianty wymagają określonej z góry docelowej liczby klastrów k , wartość ta w początkowej fazie badań ustalona została na podstawie znajomości technologii i w opisywanym przypadku wyniosła 8. Dzięki zgrupowaniu w klastry wektorów o podobnych cechach, możliwe będzie wykrycie wzorców technologicznych odpowiadających wariantom, jakie produkowane były w trakcie pracy instalacji.

Jednym z problemów związanych z przygotowaniem danych do klasteryzacji jest duża różnica w rzędach wielkości cech wektora V_{MC} . Całkowite zużycie energii wyrażone jest w pojedynczych litrach, podczas gdy czasy aktywności mierzone są w dziesiątkach tysięcy milisekund. Aby uniknąć niwelowania wagi cech ze względu na niewielkie wartości w porównaniu do innych, niezbędne jest przeprowadzenie operacji skalowania. W badanym przypadku wartość zużycia we wszystkich wektorach V_{MC} została przemnożona przez stałą wartość 10000, co zwiększyło jej rolę w realizacji algorytmu przydzielania do klastrów. Operacji przeskalowania można poddać również inne wartości cech, a mnożnik powinien być dobrany w zależności od faktycznych ich wartości, jednak w badanym przypadku nie istniała taka konieczność.

Wyniki klasteryzacji, której poddany został eksperymentalny zbiór danych, przedstawiono w tabeli 4.2.1. Zestawia ona centroidy poszczególnych cech otrzymanych klastrów: dziewięć związanych z czasami aktywacji pojedynczych urządzeń wykonawczych oraz jednej związanej z całkowitym zużyciem powietrza.

Tabela 4.2.1. – wyniki klasteryzacji zbioru eksperymentalnego dla $k = 8$

Cecha	Klaster 1	Klaster 2	Klaster 3	Klaster 4	Klaster 5	Klaster 6	Klaster 7	Klaster 8
Q20_0	1733.412	1736.118	1750.706	1747.278	1041.882	1057.222	1042.352	1082
Q20_1	9939.412	9896.353	9806.059	9726.471	918.444	910.647	908.389	908.389
Q20_2	1632.118	1612.118	1628.176	1622.333	0	0	0	0
Q20_3	6247.588	6219.294	6151.176	6061	0	0	0	0
Q20_4	1515.941	1411.118	1502.412	1396.667	311.118	282.222	310.176	322.222
Q20_5	3334.765	3379.706	3301.941	3289.944	244.176	297.278	283.412	288.833
Q20_6	307.065	315.059	307.529	308.889	0	0	0	0
Q20_7	0	0	0	0	99.353	96.056	89.647	100.167
Q21_0	770	0	3804.059	1780.222	0	1795.444	3802.353	821.611
Zużycie powietrza	50632.486	43407.054	70156.824	57606.337	30772.633	45667.318	60532.137	38270.942

Otrzymane klastry, wraz z wyznaczonymi centroidami, wskazują na wzorce technologiczne, zgodnie z którymi stanowisko DC40 realizowało warianty produkcyjne. W wyniku działania metody otrzymano osiem klastrów, odpowiadających ośmiu wzorcom potwierdzonym przez technologię: na czas trwania eksperymentu produkcja prowadzona była dla czterech wariantów pełnych, wykorzystujących niemal wszystkie urządzenia wykonawcze oraz czterech wariantów uproszczonych. Warianty uproszczone reprezentowane są przez klastry 5-8, w których centroidy dla trzech wybranych urządzeń wynoszą 0. Potwierdzeniem poprawności przydziału wektorów V_{MC} do właściwych grup jest zgodność wartości centroidów z danymi z macierzy E_{MC} .

Wspomniany wcześniej problem potencjalnej bliskości wariantów ze względu na podobne wartości cech znaczących, np. kumulacyjnego zużycia energii, możliwy jest do zaobserwowania na danych przypisanych do klastrów 2 i 6 oraz 4 i 7. Odpowiadającymi im profilami z rysunku 4.2.1. są profile 3 i 4 oraz 6 i 7. Analizując przypisanie poszczególnych cykli produkcyjnych do tych wariantów potwierdzono poprawność grupowania ich, które nie byłoby możliwe bez wprowadzenia cech wektora V_{MC} , opisujących czasy aktywności urządzeń.

Zastosowanie macierzy E_{MC} wyznaczonej w etapie przygotowania danych do analizy [i] umożliwiło przeprowadzenie skutecznego procesu klasteryzacji, któremu poddawane były kolejne wektory V_{MCn} . Liczba otrzymanych klastrów odpowiada liczbie wykrytych wzorców technologicznych, a centroidy tych klastrów wskazują na referencyjne wartości cech wektora V_{MCn} , który reprezentuje cykle produkcyjne realizowane w trakcie produkcji wariantu odpowiadającego danemu wzorcowi. Ponieważ jednak rozwój metody

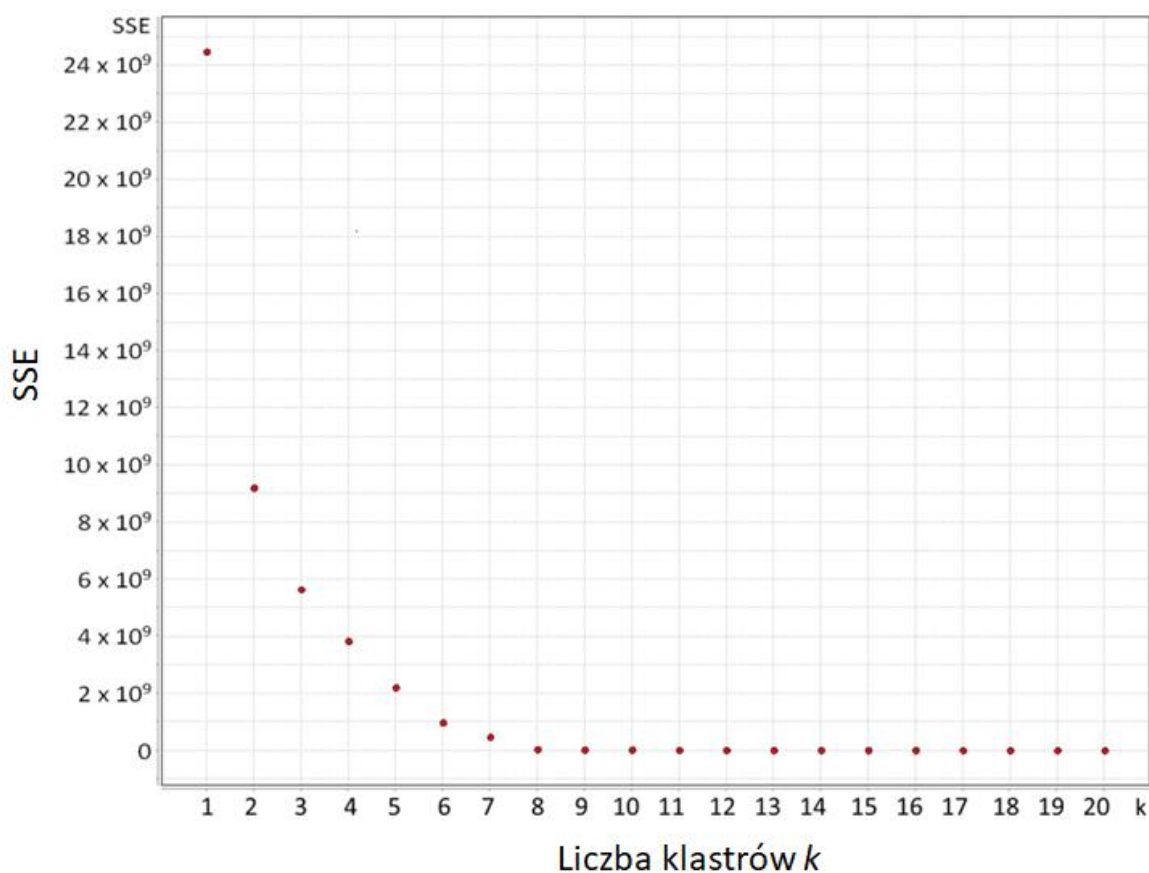
prowadzony był w stabilnych warunkach pracy instalacji, liczba produkowanych wariantów była znana i stała, co nie odzwierciedla w pełni rzeczywistego obrazu produkcji dyskretnej i wielowariantowej. W produkcji takiej warianty mogą zmieniać się dynamicznie, a algorytm detekcji powinien mieć zdolność do samodzielnego wykrywania nowych wzorców i odpowiedniego ich oznaczania. Stosowanie klasycznej metody klasteryzacji opartej o algorytm k -Means nie daje możliwości automatycznego wyznaczenia liczby k , stąd konieczna była kontynuacja badań nad automatyzacją wyznaczenia tej liczby.

4.3. Automatyzacja wyznaczenia liczby wzorców technologicznych z użyciem zmodyfikowanego algorytmu klasteryzacji k -Means

Badania nad automatyzacją wykrywania nowych wzorców technologicznych stanowiły rozwinięcie metody detekcji wzorców [j] przedstawionej na rysunku 4.1. Jak zaprezentowano w rozdziale 4.2., liczba k determinująca liczbę podziałów zbioru wejściowego zawierającego odpowiednio przygotowane wektory opisujące cykle produkcyjne, odpowiada liczbie wykrytych wzorców technologicznych. Podstawowe implementacje algorytmu k -Means wymagają podania liczby k przed rozpoczęciem operacji, a z powodu potencjalnej wielowariantowości produkcji zakładanie, że liczba ta będzie stała w trakcie całego trwania produkcji jest błędne. Dodatkowo stoi ono w opozycji do założenia, że wzorce będą wykrywane bez znajomości technologii, a więc, między innymi: wiedzy o produkowanej liczbie wariantów produkcyjnych. Z tego powodu liczba k musi być wyznaczana automatycznie, co jest trudnym i skomplikowanym zadaniem, a metody automatycznego wykrywania liczby k przedstawione w rozdziale 2.4.1. cechują się dużą złożonością uniemożliwiającą ich zastosowanie w systemach produkcyjnych, dla których wymagane może być szybkie uzyskanie rezultatów z działania po otrzymaniu nowych danych wejściowych. Ze względu na produkcję masowo-dostosowaną i krótkoseryjną, w której warianty produkcyjne pojawiać się mogą dynamicznie i być realizowane przez krótki czas, brak jest możliwości gromadzenia dużej ilości danych treningowych, niezbędnych do prawidłowego nauczenia algorytmów detekcji. Dodatkowo, specyfika analizowanych danych produkcyjnych powoduje, że wyznaczone wzorce technologiczne mogą być do siebie bardzo zbliżone, konieczne jest więc rozróżnienie pomiędzy tymi zgrupowaniami. Prace badawcze autora przeprowadzone zostały w celu

zbadania możliwości zniwelowania tych braków, a ich efektem było opracowanie nowej metody automatycznego wyznaczania liczby podziałów zbioru k dla metody k -Means.

Badania rozpoczęto od analizy metryki stosowanej do oceny jakości klasteryzacji, którą jest suma błędów kwadratowych odległości (ang. Sum of Squared Errors, SSE). W algorytmie k -Means determinuje ona jakość dystrybucji obserwacji w klastrach poprzez określenie odległości cech wektora do centroidów. Minimalizacja wyznaczania SSE w rozpatrywanym zbiorze danych prowadzi do określenia najlepszego dopasowania obiektów do poszczególnych grup. Niestety, SSE nie nadaje się do porównania jakości dystrybucji pomiędzy różnymi k . Rysunek 4.3.1. przedstawia SSE wyliczone dla zbioru eksperymentalnego zebranego podczas pracy stanowiska DC40, podzielonego na różną liczbę docelowych klastrów k . Punkty określone przez SSE tworzą monotoniczną funkcję opadającą, cechującą się spadkiem sumy błędów kwadratowych wraz ze zwiększaniem docelowej liczby grup. Wynika to z faktu, że SSE dąży do wartości 0 przy podziale zbioru na liczbę klastrów równą liczbie wszystkich obserwacji w tym zbiorze.

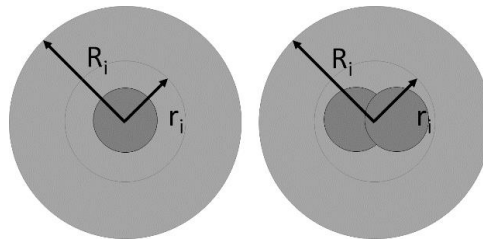


Rysunek 4.3.1. – zależność sumy błędów kwadratowych odległości od liczby k dla zbioru obserwacji pozyskanego w trakcie pracy stanowiska DC40

Stosowanie jedynie metryki SSE do oceny jakości podziału zbioru na grupy doprowadziłoby do błędnego wyznaczenia liczby klastrów. Z tego powodu konieczne jest znalezienie innej metryki oceny jakości klasteryzacji, będącej funkcją liczby k . Ponieważ dążeniem jest, aby osiągnąć naturalną liczbę klastrów, odpowiadającą rzeczywistej liczbie wzorców technologicznych realizowanych na stanowisku, poszukiwanie odpowiednich zależności może być realizowane przez porównanie wektorów wejściowych opisujących cykle produkcyjne związane z realizacją wspólnych wariantów. W analizowanych przypadkach, podobieństwa poszukiwane będą dla cech opisujących czasy aktywności pojedynczych urządzeń wykonawczych oraz całkowitej energii zużywanej w trakcie cykli produkcyjnych. Opisany w rozdziale 4.2. problem różnic w rzędach wielkości pomiędzy czasami wysterowań urządzeń w milisekundach, a zużywanymi mediami zasilającymi w litrach, również tutaj wymaga zniwelowania poprzez arbitralne przemnożenie zużycia przez wartość 10000.

Wykorzystując właściwości macierzy E_{MC} do reprezentowania poszczególnych cykli produkcyjnych w postaci odpowiednio znaczących cech oraz obserwując specyfikę pracy instalacji produkcyjnych możliwe jest przeprowadzenie analizy jakości zgrupowań w oparciu o koncentrację cech wektorów wokół pewnego teoretycznego punktu związanego z badanym wzorcem technologicznym. Jeżeli dany klaster składa się z obserwacji związanych z konkretnym wariantem produkcyjnym, obserwacje te skupione będą wokół tego punktu, zdefiniowanego jako jądro grupy. Ze względu na wielowymiarowość wektorów wejściowych, obserwacje rozmieszczone są w każdym z wymiarów tego wektora zgodnie z rozkładem normalnym, wynikającym jedynie z niestabilności parametrów procesu produkcyjnego, wpływającego np. na fluktuacje czasu pracy pojedynczych urządzeń oraz z błędów pomiarowych. Dla przypadku, gdy jeden klaster oznacza jeden wzorzec technologiczny, obserwacje skupione będą wokół pojedynczego jądra. W przypadku, gdy wzorce są do siebie zbliżone na tyle, że metoda klasteryzacji grupuje je w jeden klaster, wówczas, obserwacje skoncentrowane będą wokół więcej niż jednego jądra. Uwzględniając te właściwości, zaproponowano rozwinięcie klasteryzacji metodą k -Means o analizę skupień wokół jąder hiperkul dla potrzeb automatycznego wyznaczania liczby k . W metodzie tej każdy klaster oznaczający docelowy wzorzec technologiczny składa się z hiperkul, w których obserwacje skoncentrowane są wokół ich środków. Poszukiwanie docelowej liczby klastrów realizowane jest poprzez analizę jakości dopasowania obserwacji do jąder dla każdego

badanego przypadku k i prowadzone jest na zbiorze uczącym. Graficzna prezentacja hiperkul z pojedynczym i wielokrotnym jądrem przedstawiona została na rysunku 4.3.2.



Rysunek 4.3.2. – graficzna prezentacja hiperkul dla przypadku pojedynczego i wielokrotnego jądra

Zewnętrzny promień hiperkuli opisany został jako R_i i oznacza on maksymalną wartość sumy błędów kwadratowych odległości wyliczonej dla wszystkich obserwacji należących do danego klastra. Jednocześnie stanowi on ograniczenie przynależności tych obserwacji do i -tego klastra. Uwzględniając N -wymiarowość wektora wejściowego, promień zewnętrzny hiperkuli może zostać opisany wyrażeniem:

$$R_i = \max_{1 \leq j \leq |CL_i|} \left(\sqrt{\sum_{z=1}^N (S_z^{ij} - C_z^i)^2} \right) \quad (4)$$

gdzie:

CL_i = i -ty klaster,

$|CL_i|$ = liczba obserwacji w i -tym klastrze,

N = rozmiar wektora V_{MC}

C = centroid i -tego klastra,

S^j = j -ta próbka w i -tym klastrze

Wewnętrzny promień hiperkuli oznaczony został jako r_i i oznacza średnią wartość sumy błędów w danym klastrze. Promień ten może zostać opisany wyrażeniem:

$$r_i = \frac{1}{|CL_i|} * \sum_{j=1}^{|CL_i|} \sqrt{\sum_{z=1}^N (S_z^{ij} - C_z^i)^2} \quad (5)$$

Obliczając wartość:

$$\frac{R_i}{r_i} \quad (6)$$

możliwe jest określenie stopnia koncentracji wyznaczonego klastra, wskazującego na skupienie jego obserwacji wokół teoretycznego jądra. Obliczony iloraz mieścić się będzie w przedziale od wartości 0 do wartości 1, gdzie 0 oznacza rozlokowanie wszystkich obserwacji w jądrze hiperkuli, a 1 równość wszystkich obserwacji do maksymalnej wartości sumy błędów kwadratowych R_i . W przypadku, gdy obserwacje wewnątrz hiperkuli skupione są wokół pojedynczego jądra, wartość stopnia koncentracji dla tego klastra będzie mniejsza niż dla przypadku klastrów, które składają się z tych samych obserwacji ale które zawierają dwa lub więcej jądra. Dla przypadku pojedynczego jądra wartość r_i odwzorowuje rozkład normalny obserwacji w i -tym klastrze. Ponieważ wartość $\frac{R_i}{r_i}$ nie wskazuje w żaden sposób liczby obserwacji w klastrze, niezbędne jest jej podzielenie przez sumę elementów przyporządkowanych do grupy, a następnie zsumowanie dla wszystkich k grup.

W celu wyznaczenia funkcji jakości dopasowania obserwacji do danego klastra, możliwe jest podzielenie wyznaczonych wartości stopnia koncentracji przez liczbę uzyskanych klastrów dla sumy wszystkich otrzymanych zbiorów. Wartość ta opisana jest wyrażeniem:

$$Q(k) = \frac{1}{k} * \sum_{i=1}^k \left(\frac{R_i}{r_i} * \frac{1}{|CL_i|} \right) \quad (7)$$

Poszukiwanie docelowej liczby klastrów sprowadza się do poszukiwania wartości minimalnej funkcji $Q(k)$ dla każdego przypadku badanej liczby k . Wartość ta zdefiniowana została jako Q_1 :

$$Q_1 = \underset{k}{\operatorname{argmin}}(Q(k)) \quad (8)$$

Niestety zależność ta nie wyeliminuje przypadków odstających, czyli klastrów, składających się z pojedynczych obserwacji. Wówczas, promienie R_i oraz r_i przyjmują wartość 0, a Q_1 pozostaje nieoznaczona. Zakładając, że wszystkie występujące i badane warianty produkcyjne cechują się produkcją więcej niż jednego produktu, skutkuje to wystąpieniem więcej niż jednej obserwacji (wektora V_{MCn}) o podobnych właściwościach.

Wystąpienie przypadków odstających powinno być więc traktowane jako anomalia, a nie potencjalnie nowy wzorec do wykrycia.

Uwzględniając to założenie, końcowa wartość liczby podziałów k może zostać opisana jako $Q2$:

$$Q2 = Q1 - |LK1(k)| \quad (9)$$

gdzie:

$|LK1(k)|$ = liczba elementów odstających wykryta jako liczba klastrów składających się z jednej obserwacji, wyznaczona dla każdego badanego przypadku k .

W następnym kroku przeprowadzono walidację metody. W pierwszej kolejności przeprowadzono ją na referencyjnym zbiorze danych, zebranych w trakcie pracy stanowiska DC40. Średnie wartości poszczególnych cech wektora V_{MCn} wyznaczone dla realizacji ośmiu wariantów produkcyjnych przedstawione zostały w tabeli 4.3.1.

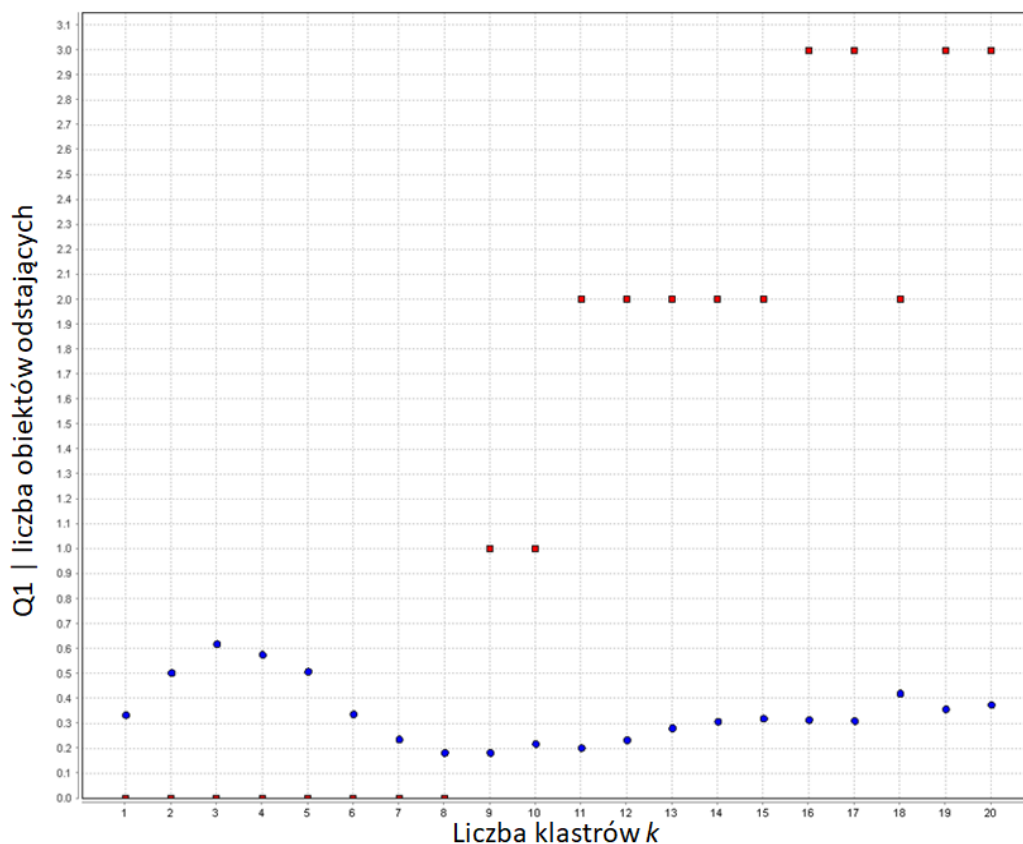
Tabela 4.3.1. – średnie wartości cech wektorów V_{MCn} dla referencyjnego zbioru danych opisującego 8 wzorcowych wariantów produkcji stanowiska DC40

	Wariant 1	Wariant 2	Wariant 3	Wariant 4	Wariant 5	Wariant 6	Wariant 7	Wariant 8
Liczba zrealizowanych cykli produkcyjnych	17	18	17	18	17	18	17	17
Q20.0	1041,9	1082	1736,1	1057,2	1733,4	1747,3	1042,4	1750,7
Q20.1	908,5	908,4	9896,4	918,4	9939,4	9726,8	910,6	9806,1
Q20.2	0	0	1612,1	0	1632,1	1622,3	0	1628,2
Q20.3	0	0	6219,3	0	6247,6	6061	0	6151,2
Q20.4	311,1	322,2	1411,1	282,2	1515,9	1396,7	310,2	1502,4
Q20.5	244,2	288,8	3379,7	297,3	3334,8	3289,9	283,4	3301,9
Q20.6	0	0	315,1	0	307,8	307,9	0	307,5
Q20.7	99,4	100,2	0	96,1	0	0	89,6	0
Q21.0	0	821,6	0	1795,4	770	1780,2	3802,4	3804,1
Całkowite zużycie powietrza [l x 10000]	30772,6	38270,9	43407,1	45667,3	50632,5	57606,3	60352,1	70156,8

Obliczenia wartości funkcji $Q1$ przeprowadzone zostały w środowisku Matlab. Wyniki otrzymane dla 20 iteracji liczby k zaprezentowane zostały w tabeli 4.3.3. oraz graficznie na rysunku 4.3.2.

Tabela 4.3.2. – wartość funkcji $Q1$ i liczba obserwacji odstających dla 20 wariantów podziału zbioru testowego zawierającego 8 wariantów produkcji

k	1	2	3	4	5	6	7	8	9	10
Q1	0,3357	0,5046	0,6202	0,5773	0,5093	0,3388	0,2381	0,1841	0,1839	0,2198
Liczba obserwacji odstających	0	0	0	0	0	0	0	0	1	1
k	11	12	13	14	15	16	17	18	19	20
Q1	0,2032	0,2354	0,2829	0,3089	0,3213	0,316	0,312	0,4216	0,359	0,3767
Liczba obserwacji odstających	2	2	2	2	2	3	3	2	3	3



Rysunek 4.3.3. – wartość funkcji $Q1$ (punkty niebieskie) i liczba obserwacji odstających (punkty czerwone) dla 20 wariantów podziału zbioru testowego zawierającego 8 wariantów produkcji

Wartość minimalna funkcji $Q1$ została osiągnięta dla liczby k równej 9, jednak jednocześnie dla takiego podziału wykryta została jedna obserwacja odstająca, stąd końcowa wartość $Q2$ oznaczająca liczbę docelowych klastrów wynosi 8. Choć wartość ta jest prawidłowa i potwierdza skuteczność zaproponowanej metody, dane użyte do jej walidacji powinny być niezaburzone i zawierać tylko wartości prawidłowe dla pracy

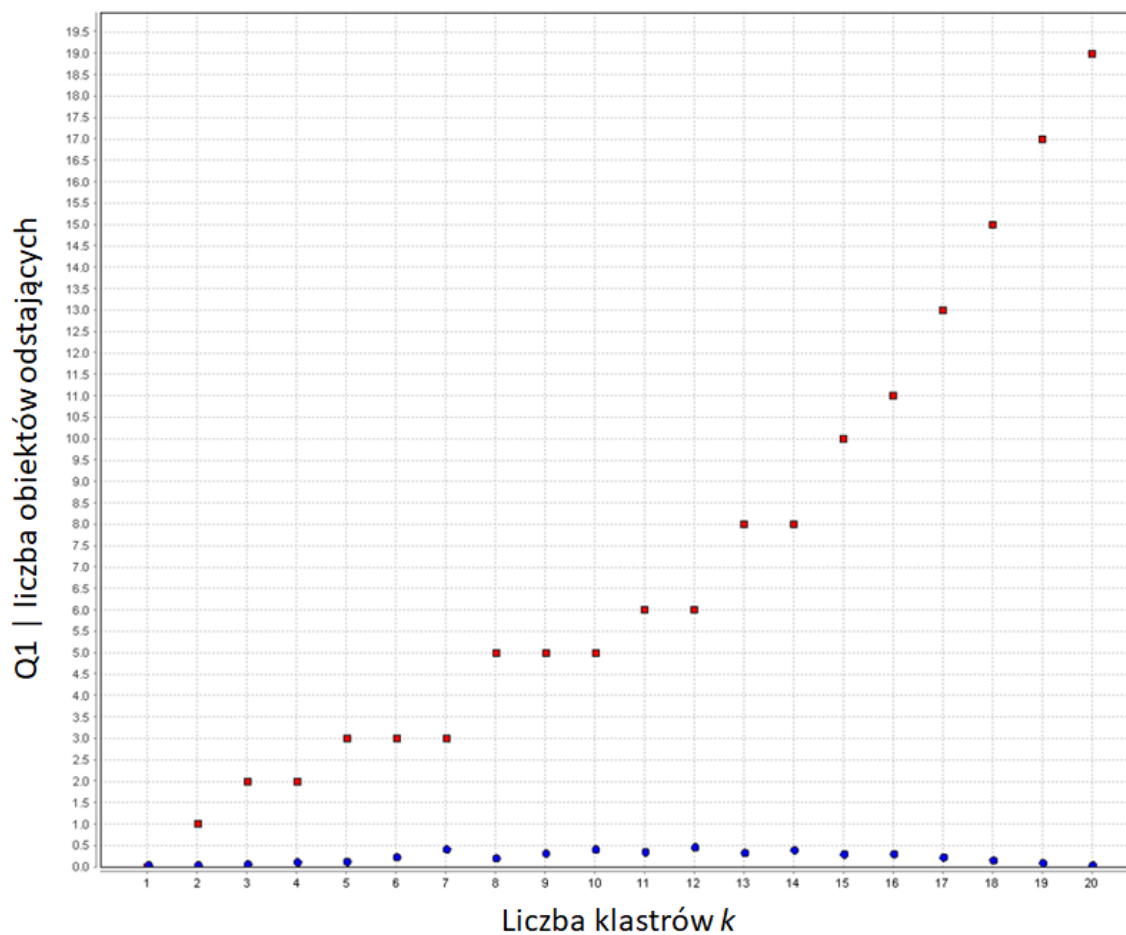
stanowiska w poszczególnych cyklach. Po analizie danych testowych stwierdzono, że w istocie w trakcie pracy stanowiska wystąpił cykl pracy cechujący się nieznacznym wydłużeniem czasu pracy jednego z urządzeń pneumatycznych, co przełożyło się na wystąpienie jednego klastra jednoelementowego.

Następnie, dalsza weryfikacja metody przeprowadzona została dla zbioru obejmującego jeden typ produkowanego wariantu i miała na celu sprawdzenie skuteczności wyznaczania liczby k dla przypadku jednolitych zbiorów danych produkcyjnych. Eksperyment objął rejestrację 21 cykli produkcyjnych stanowiska DC40, produkującego w jednym wariantcie. Otrzymane wartości $Q1$ oraz liczby obserwacji odstających dla 20 iteracji liczby k przedstawione zostały w tabeli 4.3.3. oraz na rysunku 4.3.4.

Tabela 4.3.3. – wartość funkcji $Q1$ i liczba obserwacji odstających dla 20 wariantów podziału zbioru testowego zawierającego 1 wariant produkcji

k	1	2	3	4	5	6	7	8	9	10
Q1	0,0567	0,052	0,0743	0,1243	0,1353	0,2404	0,425	0,2168	0,3278	0,421
Liczba obserwacji odstających	0	1	2	2	3	3	3	5	5	5
k	11	12	13	14	15	16	17	18	19	20
Q1	0,3601	0,4736	0,342	0,4055	0,3084	0,3125	0,2353	0,1667	0,1053	0,05
Liczba obserwacji odstających	6	6	8	8	10	11	13	15	17	19

Minimalna wartość funkcji $Q1$ została osiągnięta dla k równej 21 i wyniosła 0, jednak nie może ona zostać uznana za prawidłową ponieważ dla takiego przypadku zbiór wejściowy został podzielony na liczbę grup równą liczbie obserwacji. Z tego powodu jako prawidłowa traktowana powinna zostać wartość k równa 2, co dla obecnej w tym przypadku pojedynczej wartości odstającej oznacza wartość funkcji $Q2$ równą 1.



Rysunek 4.3.4. – wartość funkcji $Q1$ (punkty niebieskie) i liczba obserwacji odstających (punkty czerwone) dla 20 wariantów podziału zbioru testowego zawierającego 1 wariant produkcji

Trzecim eksperymentem sprawdzającym poprawność metody było wyznaczenie liczby wariantów produkcji dla zbioru zawierającego dane zaburzone, obejmujące aktywność stanowiska DC40 w trakcie różnych anomalii w pracy urządzeń. W tym celu zarejestrowano dwa zbiory danych zawierających zaburzenia: pierwszy, z 95 cyklami realizowanymi w 4 wariantach produkcyjnych i jednym cyklem z anomaliami w pracy urządzeń Q20.2 oraz Q20.3 oraz drugi, składający się ze 105 cykli produkcyjnych realizowanych dla 4 wariantów, zawierający 2 cykle z występującymi nieprawidłowościami w pracy urządzeń Q20.2 oraz Q20.3. Nieprawidłowości w pracy wspomnianych urządzeń polegały na znaczącym, minimum dziesięciokrotnym przekroczeniu typowego czasu aktywności urządzenia Q20.2 oraz aktywacji urządzenia Q20.3 pomimo tego, że w realizowanych wariantach w ogóle nie powinno ono być

używane. Otrzymane wartości $Q1$ oraz liczby obserwacji odstających dla obu tych zbiorów przedstawione zostały w tabeli 4.3.4.

Tabela 4.3.4. – wartość funkcji $Q1$ i liczba obserwacji odstających dla 20 wariantów podziału dwóch testowych zbiorów zaburzonych, zawierających po 4 warianty produkcji

k	1	2	3	4	5	6	7	8	9	10
Q1 dla pierwszego zbioru	0,3018	0,4465	0,3146	0,2402	0,09	0,262	0,259	0,3654	0,2139	0,2406
Liczba obserwacji odstających dla pierwszego zbioru	0	0	0	1	1	1	2	2	3	3
Q1 dla drugiego zbioru	0,1198	0,1811	0,2496	0,2303	0,1016	0,0996	0,1013	0,1292	0,1732	0,1856
Liczba obserwacji odstających dla drugiego zbioru	0	0	0	1	1	2	2	3	3	3
k	11	12	13	14	15	16	17	18	19	20
Q1 dla pierwszego zbioru	0,3485	0,1825	0,3767	0,2742	0,2375	0,3368	0,2979	0,3851	0,3601	0,3885
Liczba obserwacji odstających dla pierwszego zbioru	2	4	2	4	4	4	6	5	5	5
Q1 dla drugiego zbioru	0,2099	0,2411	0,2404	0,2345	0,3018	0,3046	0,2921	0,3246	0,3529	0,343
Liczba obserwacji odstających dla drugiego zbioru	3	3	3	4	3	3	4	4	4	5

Dla przypadku pierwszego zbioru funkcja $Q1$ przyjmuje wartość minimalną dla k równej 5 i jednej obserwacji odstającej. Z kolei dla drugiego zbioru funkcja $Q1$ przyjmuje wartość minimalną dla k równej 6 i dwóch obserwacji odstających. Obie wartości są prawidłowe, wykrycie odpowiedniej liczby wartości odstających potwierdza wystąpienie anomalii w trzech wykrytych cyklach produkcyjnych, a rozpoznana liczba wzorców technologicznych wynosi po 4 dla każdego ze zbiorów. Jednocześnie zaobserwować można, że różnica pomiędzy wartością funkcji $Q1$ dla docelowej liczby k oraz $Q1$ dla k równej 1 jest znacznie mniejsza dla zbioru zawierającego większą liczbę zaburzeń. Wynika to z wpływu właśnie tej liczby: każda obserwacja odstająca była unikatowa ze względu na wartość swoich cech, a przez to pojawiła się w sposób grupujący zbiór przez algorytm jako osobny wzorec produkcyjny, mający tylko jedno wystąpienie. Dla wartości k równej 1,

cykle zawierające te zaburzenia włączone zostały do jednego klastra z innymi obserwacjami, a funkcja $Q1$ osiąga tym większe ekstremum im więcej takich zaburzeń znajduje się we wspólnym klastrze z innymi obserwacjami. Zjawisko to dowodzi pewnego ograniczenia metody, polegającego na wrażliwości definiowania obserwacji jako odstającej tylko w przypadku, gdy w badanym zbiorze występuje ona jednorazowo. Jeżeli dane treningowe zawierać będą więcej niż jeden zaburzony cykl, cechujący się podobną specyfiką zaburzenia, cykle te zostaną zgrupowane w dodatkowy, nieprawidłowy wzorzec technologiczny, a cały zbiór wejściowy może zostać zgrupowany w jednym wzorcu o dużym rozrzucie obserwacji, co z kolei doprowadzić może do nieprawidłowego wyniku metody. Istotnym jest więc, aby dane służące do nauki wzorców były możliwie najbardziej prawidłowe oraz przede wszystkim, aby nie zawierały wielokrotnych zaburzeń tego samego typu.

Automatyzacja liczby k definiującej liczbę podziałów zbioru na wzorce technologiczne przeprowadzona została z użyciem zmodyfikowanej metody klasteryzacji k -Means. Metoda ta wykorzystuje analizę skupień cech obserwacji wokół teoretycznego jądra wzorca, wyznaczanego w kolejnych iteracjach podziału zbioru na różną liczbę k . Dzięki zbudowaniu funkcji jakości dopasowania do danego klastra możliwe jest określenie najlepszego podziału zbioru na grupy docelowe. Jednocześnie, nieuwzględnienie w końcowym podziale grup formowanych przez pojedyncze obserwacje zapewnia detekcję faktycznych wzorców, składających się z co najmniej dwóch cykli produkcyjnych. Niestety ta sama właściwość powoduje, że metoda jest potencjalnie wrażliwa na grupowanie w postaci nowego wzorca podobnych do siebie cykli zaburzonych. Stąd konieczne jest, aby na czas detekcji wzorców technologicznych rejestrowane dane procesowe nie zawierały informacji o anomaliach w pracy badanej instalacji.

4.4. Podsumowanie

Zaproponowana metoda detekcji wzorców produkcyjnych ogranicza złożony problem uogólnionego przypadku matematycznego badania zużycia energii dla różnych, nieznanymi wariantów produkcyjnych poprzez porównanie obserwowanych cykli produkcyjnych opisanych w postaci wektorów V_{MCn} i zachowań układu wykonawczego opisanego w postaci macierzy zużycie energii E_{MC} .

Agregacja danych do postaci wektorów V_{MCn} opisujących poszczególne cykle znacząco redukuje problem ilości danych poddawanych dalszej analizie oraz umożliwia

ich wstępną filtrację do postaci opisującej tylko pracę urządzeń mających istotny wpływ na dany cykl produkcyjny.

Zastosowanie metody k -Means do grupowania obserwacji opisanych w postaci wektorów V_{MCn} cechuje się prostotą implementacji, łatwością stosowania dla wystąpienia nowych wzorców technologicznych i umożliwia szybkie określenie referencyjnych wartości dla nowych obserwacji, przypisywanych do klastrów poprzez automatyczne wyznaczanie centroidów.

Automatyzacja wyznaczania liczby podziałów zapewnia możliwość zgrupowania cykli produkcyjnych bez konieczności posiadania złożonej wiedzy na temat technologii wykonania instalacji produkcyjnej lub odgórnej informacji o aktualnie produkowanym wariantcie. Jest to szczególnie istotne dla produkcji adaptacyjnej, w której występowanie nowych wariantów może być częste i niekoniecznie zarządzane przez nadrzędny system kontroli produkcji.

Dodatkowo, metoda detekcji wzorców nie wymaga w żaden sposób poszeregowania lub zachowania porządku w obserwacjach zbioru uczącego. Powoduje to, że grupowanie i wyznaczanie wartości referencyjnych opisujących poszczególne profile produkcyjne może być prowadzone niezależnie od stosowanej technologii czy wiedzy na temat planów produkcyjnych.

Ograniczeniem metody automatyzacji wyznaczania liczby k jest jej wrażliwość na wektory V_{MCn} zawierające zaburzone wartości cech, w szczególności opisujące wielokrotne występowanie podobnych do siebie anomalii. Ważne jest więc, aby dane treningowe stosowane do nauki nowych wzorców zawierały w miarę możliwości dane niezaburzone. Jednocześnie metoda ta cechuje się odpornością na drobne nieprawidłowości w realizacji produkcji, wynikające z niedoskonałości wykonania instalacji lub zależności czasowych wynikających np. z kombinacyjnego sterowania urządzeniami wykonawczymi. Drugim ograniczeniem może być potencjalny wpływ zbyt wielu analizowanych jednocześnie sygnałów sterujących urządzeniami wykonawczymi na jakość wykrywanych wzorców technologicznych. Założeniem dla prac badawczych było zastosowanie tylko jednego źródła pomiaru energii, a obraz całkowitego jej zużycia może być zaburzony przez zależność od wszystkich odbiorników w instalacji. Zjawisko to będzie szczególnie widoczne w złożonych układach kombinacyjnych, np. rozległych instalacjach wyposażonych w wiele podobszarów produkcyjnych. Z tego powodu może zaistnieć

potrzeba fragmentacji stacji ze względu na pomiary zużycia energii i prowadzenie detekcji wzorców z dokładnością do wydzielonych podobszarów instalacji.

Prace badawcze skoncentrowane wokół zagadnień opisanych w tym rozdziale zostały przedstawione i opublikowane w czasopiśmie naukowych i podczas konferencji:

- Cupek, R., Ziebinski, A., Zonenberg, D., & Drewniak, M. (2018). Determination of the machine energy consumption profiles in the mass-customised manufacturing. *International Journal of Computer Integrated Manufacturing*, 31(6), 537-561.
- Cupek, R., Drewniak, M., & Zonenberg, D. (2014, June). Online energy efficiency assessment in serial production-statistical and data mining approaches. In *2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE)* (pp. 189-194). IEEE.
- Cupek, R., Ziębiński, A., Drewniak, M., & Fojcik, M. (2018, March). Improving KPI based performance analysis in discrete, multi-variant production. In *Asian Conference on Intelligent Information and Database Systems* (pp. 661-673). Springer, Cham.
- Cupek, R., Drewniak, M., & Steclik, T. (2021, June). Data Preprocessing, Aggregation and Clustering for Agile Manufacturing Based on Automated Guided Vehicles. In *International Conference on Computational Science* (pp. 458-470). Springer, Cham.
- Cupek, R., Duda, J., Zonenberg, D., Chłopaś, Ł., Dziędziel, G., & Drewniak, M. (2017, September). Data mining techniques for energy efficiency analysis of discrete production lines. In *International Conference on Computational Collective Intelligence* (pp. 292-301). Springer, Cham.
- Cupek, R., Ziębiński, A., Drewniak, M., & Fojcik, M. (2018, March). Estimation of the Number of Energy Consumption Profiles in the Case of Discreet Multi-variant Production. In *Asian Conference on Intelligent Information and Database Systems* (pp. 674-684). Springer, Cham..

5. METODA AUTOMATYCZNEGO WYKRYWANIA ANOMALII W PRACY URZĄDZEŃ

Niniejszy rozdział opisuje badania nad rozwiązaniem dwóch zagadnień przedstawionych w rozdziale 1.: wykorzystania mechanizmów eksploracji danych do poprawy jakości i efektywności procesów produkcyjnych w produkcji dyskretniej oraz zastosowania metod analitycznych bazujących na danych zagregowanych do wykrywania nieprawidłowości w pracy instalacji produkcyjnych.

Posiadanie informacji o wzorcach technologicznych pozwala na wykorzystanie jej do analizy pracy instalacji poprzez bieżącą ocenę przynależności danego cyklu produkcyjnego do konkretnego wzorca, a w przypadku braku dopasowania, stwierdzenia, jakie parametry produkcyjne prowadzą do niezgodności oraz czy niezgodność ta może być spowodowana anomalią w pracy stanowiska produkcyjnego. Niezgodność w parametrach pojedynczego cyklu produkcyjnego może wskazywać tak na anomalię, jak na błędy w pomiarze lub przygotowaniu agregatów z mierzonych wielkości fizycznych, natomiast występowanie zaburzeń powtarzających się w czasie może sugerować uszkodzenie lub degradację któregoś z elementów maszyny. Mając możliwość wykrycia takiego zjawiska, możliwe jest podjęcie działań naprawczych mających na celu wyeliminowanie usterki, która bez podjętych czynności zapobiegawczych może doprowadzić do poważnej awarii stanowiska, a w efekcie przestoju i znaczącego wzrostu kosztów, wynikającego z braku pracy i wymiany uszkodzonych komponentów na nowe.

5.1. Dopasowanie bieżących obserwacji do wyznaczonych wzorców technologicznych

Dotychczasowe operacje mające na celu wyznaczenie liczby i budowanie wzorców technologicznych prowadzone były na zamkniętym zestawie danych i realizowane w postaci analizy post-procesowej. Wykorzystanie wyników tych algorytmów do budowania bieżącej informacji o procesie wymaga jednak albo analizy danych bieżących, napływających z badanej instalacji w czasie rzeczywistym albo drugiego, zamkniętego i analizowanego post-procesowo zbioru danych. Zarówno w pierwszym jak i drugim wariancie konieczne jest stosowanie modelu wzorców, a w ramach bieżącego jego stosowania, realizowanie kolejnych zadań analitycznych, m.in. polegających na detekcji i klasyfikacji anomalii. Podczas badań autor skupił się na rozwoju metody na zbiorach

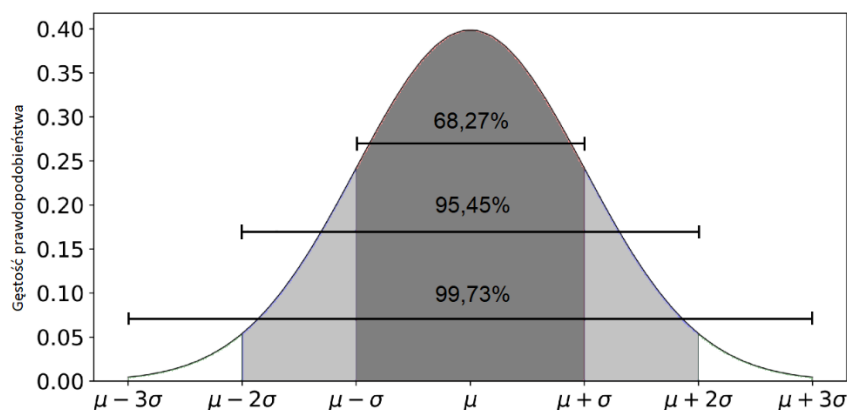
danych rejestrowanych podczas pracy stanowiska, poddawanego analizie w trybie postprocesowym.

Ponieważ rozważana pula zagadnień dotyczyła detekcji nieprawidłowości w pracy układu wykonawczego, przygotowano nowy scenariusz pracy eksperymentalnego stanowiska DC40 zakładający produkcję ośmiu wariantów. W pierwszej fazie badań, w celu wykrycia i zbudowania wzorców technologicznych, zarejestrowano produkcję 117 cykli niezaburzonych, tj. takich, które nie zawierały anomalii w postaci wycieków powietrza i wydłużonych czasów aktywności poszczególnych urządzeń wykonawczych. Zarejestrowany zbiór poddany został procesowi klasteryzacji opisanemu w rozdziale 4, a wykryte wzorce wraz z wartościami centroidów poszczególnych cech przedstawione zostały w tabeli 5.1.1.

Tabela 5.1.1. - centroidy cech referencyjnych dla ośmiu wykrytych wzorców produkcyjnych

	Q20_0	Q20_1	Q20_2	Q20_3	Q20_4	Q20_5	Q20_6	Q20_7	Q21_0	Zużycie powietrza
Wzorzec 1	1784.79	9698.64	1653.79	6246.64	1442.43	3440.64	298.21	0.00	3789.07	73317.52
Wzorzec 2	1107.75	856.5	0.00	0.00	300.75	308.88	0.00	97.38	3804.88	62105.71
Wzorzec 3	1105.44	858.13	0.00	0.00	307.50	274.19	0.00	99.56	1787.44	47071.53
Wzorzec 4	1806.42	9801.67	1673.58	6314.42	1322.33	3557.00	300.33	0.00	0.00	45846.35
Wzorzec 5	1096.6	850.27	0.00	0.00	307.40	310.13	0.00	61.60	798.07	39820.31
Wzorzec 6	1099.00	860.31	1665.00	0.00	309.50	304.44	0.00	89.69	0.00	32176.51
Wzorzec 7	1771.79	10066.29	1665	6596.57	1531.00	3567.29	305.00	0.00	782.50	53302.18
Wzorzec 8	1781.21	9873	1661.86	6402.43	1420.71	3487.71	324.00	0.00	1797.07	60181.36

Wyznaczone centroidy wskazują na wartość wzorcową cech wektorów V_{MCn} . Brakującym elementem umożliwiającym dopasowanie bieżących obserwacji (cykli produkcyjnych) do wzorców i wykorzystanie ich w dalszych analizach jest określenie przedziału przynależności do wzorca na podstawie stopnia dopasowania do poszczególnych jego centroidów. W tym celu zastosować można podejście znane ze statystyki, polegające na zastosowaniu rozkładu normalnego, dla którego zdefiniowanie progu na poziomie trzykrotności odchylenia standardowego od wartości środkowej pokrywa 99.73% obiektów przynależących do grupy [127]. Graficzna prezentacja metody przedstawiona została na rysunku 5.1.1.



Rysunek 5.1.1. – rozkład normalny gęstości prawdopodobieństwa i stopień jego pokrycia przez wielokrotność odchylenia standardowego od wartości średniej

W przypadku wzorców technologicznych dla każdego centroidu konieczne jest więc wyznaczenie odchylenia standardowego wyrażonego w postaci:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{|S|} (S_i - c)^2}{|S|}} \quad (10)$$

gdzie:

$|S|$ = liczba obserwacji

S_i = i-ta obserwacja

c = centroid

Następnie możliwe jest zbudowanie przydziału przynależności obserwacji do grupy jako trzykrotności odchylenia standardowego na plus i na minus od centroidu grupy:

$$c_i - 3 * \sigma_i \leq x_i \leq c_i + 3 * \sigma_i \quad (11)$$

gdzie:

x_i = i-ta cecha wektora V_{MCn}

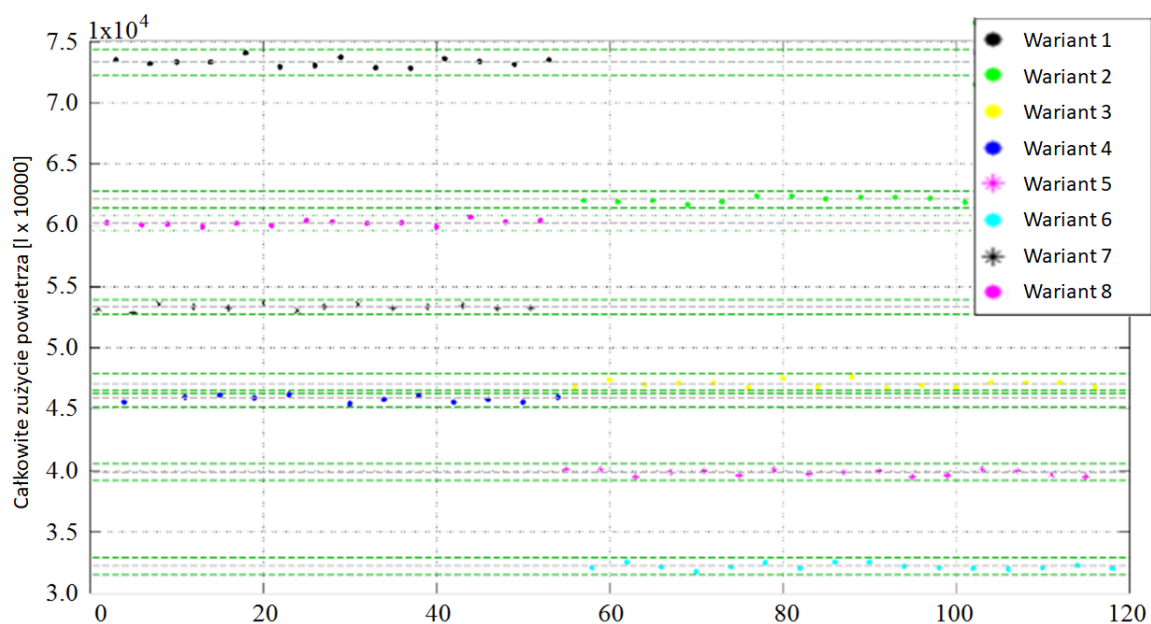
c_i = i-ta cecha centroidu

Wartości odchyłeń standardowych dla centroidów ośmiu wyznaczonych wzorców przedstawione zostały w tabeli 5.1.2.

Tabela 5.1.2. – odchylenia standardowe dla cech centroidów wyznaczonych wzorców technologicznych

	Q20_0	Q20_1	Q20_2	Q20_3	Q20_4	Q20_5	Q20_6	Q20_7	Q21_0	Zużycie powietrza
Wzorzec 1	53.146	245.625	37.031	252.56	214.7	111.8	50.43	0.00	52.344	350.966
Wzorzec 2	37.094	49.3126	0.00	0.00	47.757	48.09	0.00	47.36	47.504	231.231
Wzorzec 3	36.759	39.7859	0.00	0.00	41.47	55.181	0.00	40.01	54.889	275.057
Wzorzec 4	54.082	257.265	51.21	240.72	138.68	260.34	50.33	0.00	0.00	241.277
Wzorzec 5	26.199	44.8959	0.00	0.00	46.727	41.059	0.00	51.09	49.741	225.753
Wzorzec 6	25.859	52.2771	0.00	0.00	47.083	40.812	0.00	45.31	0.00	235.562
Wzorzec 7	62.503	439.157	53.922	410.44	373.25	208.24	41.51	0.00	44.76	202.318
Wzorzec 8	45.455	368.651	45.745	353.71	139.38	114.85	6.598	0.00	57.926	211.253

Graficzna reprezentacja granic dopasowania do klastra za pomocą trzykrotności odchylenia standardowego do centroidów wzorców przedstawiona została na rysunku 5.1.2.



Rysunek 5.1.2. – zbiór 117 cykli niezaburzonych służący do wyznaczenia 8 wzorców technologicznych wraz z granicznymi odchyleniami od centroidów

Na wykresie przedstawiono tylko jeden wymiar analizowanych wektorów V_{MCn} , tj. całkowite zużycie powietrza w trakcie trwania cyklu. Oś X przedstawia kolejne wykonywane cykle produkcyjne. Należy jednocześnie pamiętać, że analizie poddane zostało dodatkowe 9 cech wektorów odpowiadających czasom aktywności poszczególnych urządzeń wykonawczych. Na wykresie widoczny jest sposób prowadzenia procesu produkcyjnego, w którym najpierw w pierwszej połowie eksperymentu realizowano

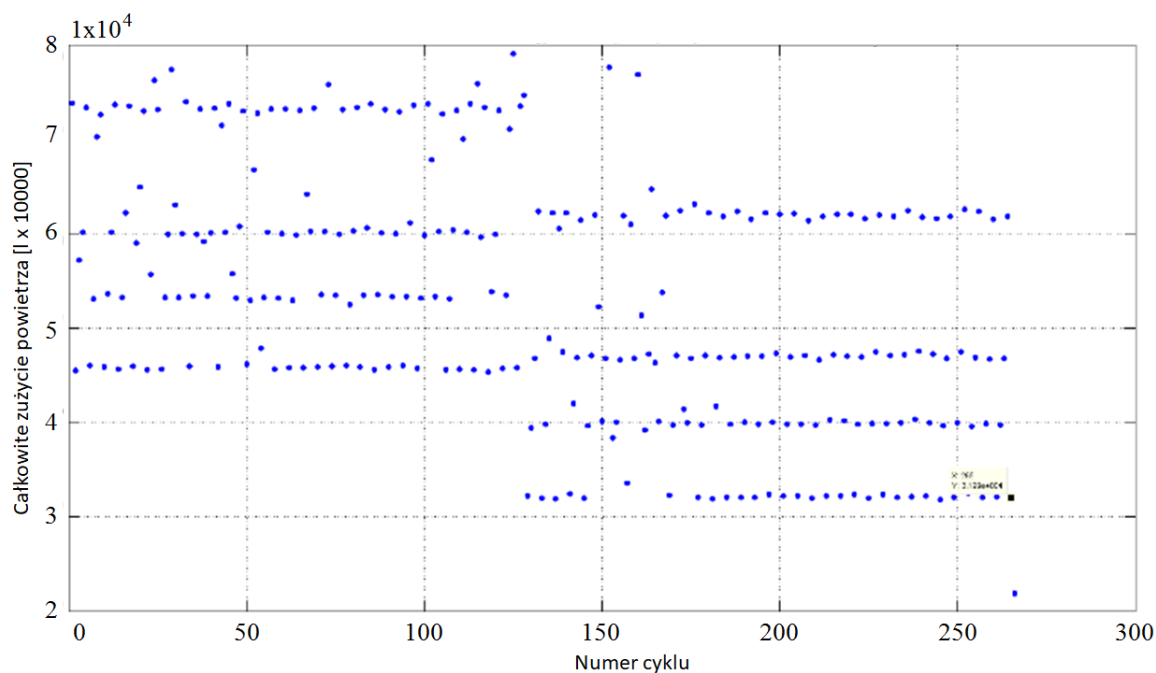
naprzemiennie cztery warianty produkcyjne, a następnie w drugiej połowie kolejne naprzemiennie cztery. Przerwane linie koloru zielonego odpowiadają przedstawionym granicom trzykrotności odchylenia standardowego od centroidu, z kolei przerywana linia szara prezentuje wartość centroidu wzorca.

Dzięki metodzie dopasowania obserwacji do wzorców technologicznych możliwe jest dalsze prowadzenie analiz, w tym wykrywanie anomalii w pracy urządzeń, przejawiające się jako niedopasowanie jednej z cech bieżących obserwacji do wyznaczonych granic dla wzorców.

5.2. Detekcja i klasyfikacja anomalii w pracy urządzeń

Posiadając metrykę dopasowania obserwacji do wzorca, przeprowadzono drugą fazę eksperymentu, polegającą na zarejestrowaniu nowego zestawu danych procesowych podczas pracy na tych samych wariantach produkcyjnych. W fazie tej wprowadzono jednak zaburzenia w pracy instalacji. Pierwszym typem zaburzenia było programowe wydłużenie czasu aktywności wybranych siłowników pneumatycznych w losowo wybranych cyklach produkcyjnych. Typ ten miał na celu zasymulowanie uszkodzenia urządzenia wykonawczego, powodujące wydłużenie właściwego czasu pracy, ale nie blokujące jej wykonania, np. odkształcenie tłoka poruszającego się wewnątrz komory siłownika pneumatycznego, przez które następuje wydłużenie czasu wykonywania ruchu lub zatarcie łożyska, generujące większe opory toczenia i spowalniające ruch obrotowy. Drugim typem zaburzenia było wprowadzenie dodatkowego, nieplanowanego zużycia energii pochłanianej przez układ poprzez dodanie polecenia otwarcia zaworu na jednym z niemonitorowanych wyjść pneumatycznych, symulujące typowy ciągły wyciek powietrza z układu, przejawiający się ciągłym, nadmiarowym i chwilowym zużyciem. Anomalia ta, oprócz zwiększonego zużycia, przejawiać się może również wydłużonym czasem aktywności urządzenia. Dla przypadku pneumatycznego będzie to wyciek, mający miejsce w torze doprowadzającym powietrze (np. uszkodzony przewód pneumatyczny na wejściu siłownika, powodujący wyciek w chwili wysterowania, ale będący jednocześnie na tyle nieznaczny, że umożliwi wykonanie ruchu tłoka, wprowadzając jednak, poprzez spadek ciśnienia w układzie, wydłużenie czasu pracy). Z kolei dla przypadku elektrycznego może to być zestaw dowolnych nieprawidłowych warunków pracy, powodujący niepożądaną aktywację urządzenia (np. uszkodzenie termostatu w szafie sterowniczej, powodujące ciągłą pracę układu chłodzącego). Ponieważ symulowane

zjawiska mogą występować niezależnie od standardowego cyklu pracy i niekoniecznie być związane z pracą jakiegokolwiek urządzenia wykonawczego, opracowano dodatkowe procedury sterujące na poziomie sterownika PLC, wprowadzające zaburzenia w sposób losowy. Dodatkowy, niemonitorowany zawór otwierany był na losowo wybrany czas od 500 do 6500 milisekund w przerwach nie krótszych niż jeden pełny cykl produkcyjny stanowiska i nie dłuższych niż 125 sekund. Typowe czasy cykli pracy przy produkowanych wariantach trwały około 14 sekund, anomalie generowane więc były od co dwóch do co dziewięć cykli. W przygotowanym w ten sposób scenariuszu testowym zarejestrowano 266 cykli zaburzonych anomaliami, w których realizowano 8 wariantów produkcyjnych, odpowiadających wariantom, dla których przeprowadzono proces nauki modelu z pierwszej fazy eksperymentu. Całkowite zużycie powietrza dla każdego z cykli produkcyjnych zarejestrowanych w zbiorze przedstawione zostało na rysunku 5.2.1.



Rysunek 5.2.1. – zużycie powietrza w poszczególnych cyklach w zbiorze z anomaliami w pracy

Wykres prezentuje liczne odchylenia zużycia powietrza od poziomych uformowań reprezentujących poszczególne warianty produkcyjne, przedstawione na rysunku 5.1.2. Wskazują one na zaburzenia w pracy instalacji, wynikające ze zwiększonego zużycia powietrza w poszczególnych cyklach produkcyjnych. Pierwszym krokiem w analizie zbioru jest przyporządkowanie poszczególnych cykli produkcyjnych do wzorców

technologicznych z wykorzystaniem mechanizmu opisanego we wcześniejszym rozdziale. Klasyfikację zrealizowano poprzez określenie stopnia dopasowania poszczególnych cech do przedziałów wokół centroidów: w przypadku, gdy wszystkie analizowane cechy leżały w granicach określonych przez wzorzec, cykl klasyfikowany był jako odpowiadający temu wzorcowi wariant produkcyjny. W przypadku, gdy co najmniej jedna cecha wykraczała poza przedział wzorca, cykl klasyfikowany był jako anomalia. Jednocześnie niezbędne było określenie stopnia dopasowania do danego wzorca tak, aby stwierdzić do jakiego wariantu produkcyjnego powinien zostać sklasyfikowany dany cykl. Do tego celu wybrano odległość liczoną pomiędzy wszystkimi cechami danej obserwacji do centroidów wariantu, liczoną osobno dla każdego wariantu:

$$d = \sqrt{\sum_{i=1}^n (c_i - S_i)^2} \quad (12)$$

gdzie:

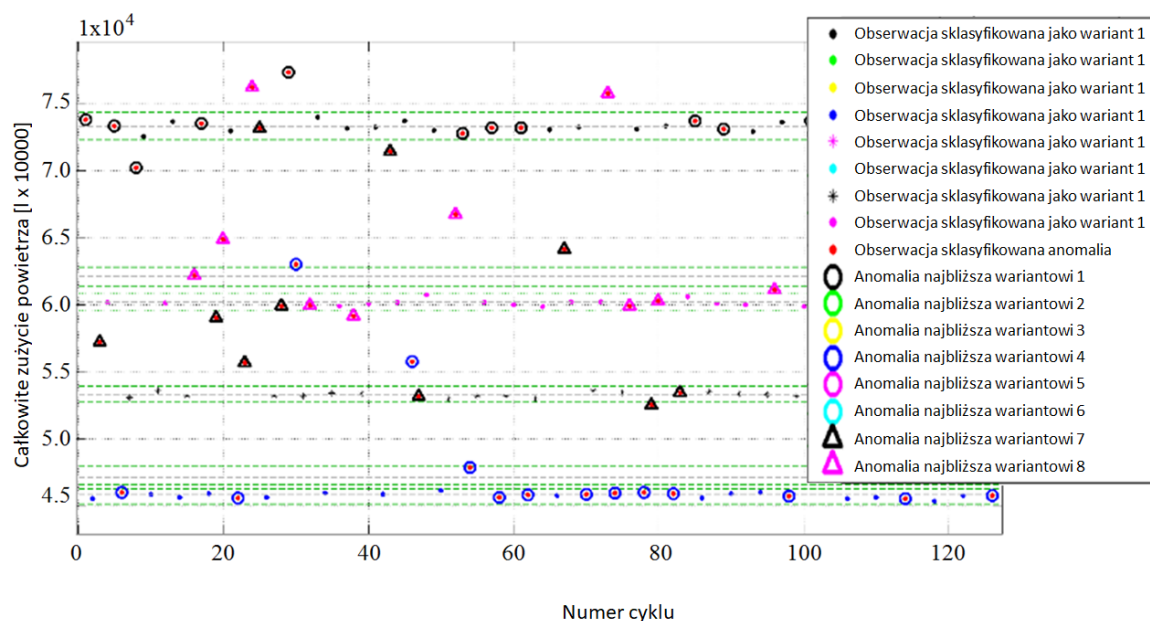
n = rozmiar wektora V_{MCn} ,

c_i = i -ta cecha centroidu danego wzorca,

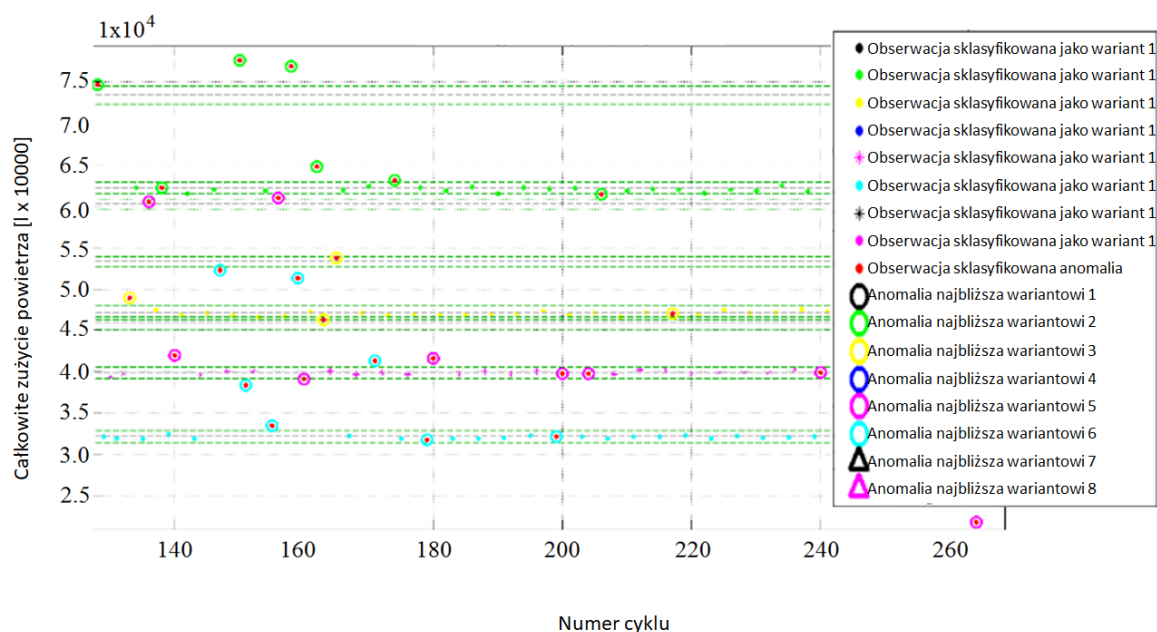
S_i = i -ta cecha analizowanego wektora V_{MCn} .

Wzorzec, dla którego wyznaczona odległość była najmniejsza, oznacza wariant, który jest najbardziej zbliżony do analizowanego cyklu z anomalią.

Po operacji klasyfikacji poszczególne obserwacje dopasowane zostały do poszczególnych wzorców. Graficzna prezentacja dopasowania przedstawiona została na rysunkach 5.2.2. i 5.2.3. Rysunki przedstawiają dwie części analizowanego zbioru wejściowego: pierwszy prezentuje okres około 130 cykli z produkcji pierwszych czterech wariantów, a drugi z kolejnych czterech. Jak wcześniej, kolorem zielonym zaznaczono granice przynależności do danego wariantu, a kolorem szarym wartość centroidu prezentowanej cechy.



Rysunek 5.2.2. – wyniki klasyfikacji obserwacji z pierwszej części zbioru wejściowego



Rysunek 5.2.3. – wyniki klasyfikacji obserwacji z drugiej części zbioru wejściowego

Obserwacje dopasowane do poszczególnych wzorców oznaczone zostały kolorami zgodnymi z opisami legendy, z kolei anomalie oznaczone zostały kolorem czerwonym. Dodatkowo rysunki prezentują wyniki dopasowania anomalii do wzorców: kształty o odpowiednich kolorach odpowiadają wzorcom, które zostały sklasyfikowane jako najbliższe zgodnie z metryką odległości cech obserwacji do cech centroidu wzorca. Podobnie jak w przypadku wcześniejszych reprezentacji graficznych zbioru danych,

również tutaj wykresy prezentują jedynie jeden wymiar analizowanych wektorów V_{MCn} , podczas gdy pozostałe 9 wymiarów jest niewidocznych.

Choć zaprezentowane wyniki dotyczą jedynie jednej cechy analizowanych obserwacji, jaką jest całkowite zużycie powietrza na cykl produkcyjny, specyfika zachowania pozostałych cech, ze względu na sposób agregacji danych, jest taka sama jak w analizowanym przypadku całkowitego zużycia powietrza na cykl. Niektóre z anomalii są łatwe w identyfikacji: brak wystąpienia wartości zużycia całkowitego wewnątrz jakiegokolwiek wzorca produkcyjnego i ulokowanie wartości w miejscach nieprzypisanych do żadnego z nich, sugeruje wystąpienie anomalii powodującej zwiększone lub zmniejszone zużycie całkowite, zwykle zwiększone poprzez wystąpienie wycieku. Znacznie trudniejszym przypadkiem, niemożliwym do wykrycia tylko za pomocą analizy graficznej, jest wystąpienie zwiększonego zużycia powietrza i uplasowanie go wewnątrz innego wzorca technologicznego, niż tego, do którego cykl produkcyjny faktycznie powinien przynależeć. Z tego powodu konieczne jest wykorzystanie wypracowanej wcześniej informacji o wzorcu technologicznym, który jest najbliższy danemu cyklowi produkcyjnemu z anomalią.

Szczegółowa analiza obserwacji z anomaliami, dopasowanych do poszczególnych wzorców, przeprowadzona została poprzez selekcję kilku wybranych cykli produkcyjnych. Wybrane obserwacje dotyczyły cyklu numer 17, 126, 159 i 174, po dwa z każdej grupy wariantów produkowanych podczas drugiej fazy eksperymentu. Wektory V_{MCn} odpowiadające tym obserwacjom przedstawione zostały w tabeli 5.2.1. Oprócz zestawienia cech w tabeli ujęto również ich przypisanie do najbliższych im wzorców technologicznych. Cechy wektorów, które nie mieściły się w granicach narzuconych przez klasyfikację metodą trzykrotności odchylenia standardowego oznaczone zostały kolorem czerwonym.

Tabela 5.2.1. – wybrane przypadki sklasyfikowanych anomalii

Numer cyklu	Q20_0	Q20_1	Q20_2	Q20_3	Q20_4	Q20_5	Q20_6	Q20_7	Q21_0	Zużycie powietrza	Wzorec technologiczny
17	1858	10801	1746	7308	1639	4471	330	0	3718	73535.2	1
126	1872	59277	1742	55694	50650	3717	321	0	0	45761.72	4
159	1086	760	0	0	329	324	0	109	0	51328.13	6
174	1086	870	0	0	324	221	0	123	3820	63085.94	2

Pierwszym z analizowanych cykli jest cykl numer 17, który przyporządkowany został do wzorca technologicznego numer 1. Cechował się anomalią polegającą na nieznacznym przekroczeniu dopuszczalnego czasu pracy urządzeń PODNOŚNIK W DÓŁ, WYPYCHACZ W TYŁ i STOP W DÓŁ podczas, gdy całkowite zużycie sprężonego powietrza mieściło się w granicach określonych wzorcem. Oznacza to, że w trakcie realizacji cyklu produkcyjnego nastąpiło zdarzenie powodujące zatrzymanie pracy tych urządzeń lub wywołujące ich wydłużony ruch. Śledząc zachowanie stanowiska od strony technologii w istocie doszło w tym cyklu do nieplanowanego wymuszenia anomalii, polegającym na kilkusekundowym zablokowaniu mechanicznym produktu na przejeździe pomiędzy dwoma urządzeniami wykonawczymi STOP i PODNOŚNIK, co w efekcie wywołało odpowiedź stanowiska w postaci przedłużenia czasu wysterowania urządzenia STOP W DÓŁ tak, aby nie doszło do uszkodzenia transportowanego produktu przy powrocie STOP do pozycji początkowej oraz PODNOŚNIK W DÓŁ i WYPYCHACZ W TYŁ, które z powodu przetrzymania produktu musiały pozostać w tych pozycjach aż do momentu pojawienia się produktu w dalszych krokach procesu. Wszystkie trzy urządzenia o przekroczonym czasie aktywności są siłownikami pneumatycznymi, które do osiągnięcia krańcowej pozycji (przy założeniu szczelności komory i przewodów doprowadzających powietrze) wymagają skończonej ilości powietrza i ich dalsze wysterowywanie nie powoduje dalszego przyrostu zużycia energii, stąd brak wpływu pracy tych urządzeń na całkowitą wartość zużycia powietrza w cyklu. Przykład podtrzymania wysterowania tak, jak dla przypadku urządzeń PODNOŚNIK i WYPYCHACZ, jest często stosowanym zabiegiem zapobiegającym potencjalnemu rozpozycjonowaniu siłowników wykonawczych, np. pod wpływem działania siły grawitacyjnej na masę umieszczoną na końcu tłoka.

Drugim z analizowanych cykli, 126, sklasyfikowany został jako najbliższy wzorcowi numer 4. Przypadek ten jest bardzo zbliżony do cyklu numer 17, z różnicami polegającymi na przyblokowaniu produktu na przejeździe z urządzenia PRE-STOP na STOP oraz

wydłużeniu czasu ruchów STOP W DÓŁ, PODNOŚNIK W DÓŁ oraz WYPYCHACZ W TYŁ do niemal minuty, a więc nawet do 600% wartości referencyjnej czasu trwania tych ruchów. Tak duża anomalia wynika najprawdopodobniej z modyfikacji założeń dla procesu w trakcie trwania eksperymentu: cykl ten jest ostatnim z serii produkcji pierwszych czterech wariantów i w jego trakcie doszło do załadowania nowych parametrów pracy dla stanowiska, co spowodowało chwilowe zatrzymanie jego pracy.

Trzeci z cykli, 159, wskazuje na typowy, ciągły wyciek podczas pracy jednego z urządzeń wykonawczych. Zaobserwować można w nim wzrost zużycia powietrza o około 70% w stosunku do wartości typowej dla wzorca numer 6. Ponieważ czas pracy urządzeń nie uległ wydłużeniu, wnioskować można, że w rzeczywistych warunkach produkcyjnych wyciek związany byłby z pracą urządzenia pobierającego energię w sposób ciągły, np. pistoletu do powietrza lub chwytaka podciśnieniowego. Brak jest dodatkowych wskazań umożliwiających wskazanie potencjalnego miejsca lub urządzenia, w którym wystąpiła anomalia.

Cykl oznaczony numerem 174 przypisany został do wzorca numer 2 i cechuje się nieznacznym przekroczeniem całkowitego zużycia powietrza. W cyklu tym najprawdopodobniej niemonitorowany zawór zadziałał na bardzo krótki czas, co wykryte zostało jako anomalia ze względu na zwiększone zużycie powietrza, bez obserwowanego wydłużonego czasu pracy jakiegokolwiek z urządzeń. Nieprawidłowość w zużyciu była bardzo niewielka i wyniosła zaledwie 0.1 litra powietrza.

5.3. Podsumowanie

Przeprowadzona analiza wybranych cykli produkcyjnych miała na celu walidację metody detekcji anomalii ze względu na niedopasowanie do wzorców technologicznych oraz klasyfikację ich ze względu na czasy aktywności urządzeń wykonawczych lub całkowitą zużytą energię podczas cyklu produkcyjnego. Zastosowane do tego celu mechanizmy statystyczne cechują się prostotą i szybkością działania, co, podobnie jak w przypadku wyznaczania wartości zagregowanych, umożliwia ich przeniesienie i uruchomienie nawet na poziomie kontrolerów procesu. Dzięki temu klasyfikacja i wykrywanie anomalii mogą być prowadzone w trybie czasu rzeczywistego, co nie było zaprezentowane w żadnej z pozycji literaturowych przedstawionych w rozdziale 2.7. Otwiera to możliwości szybkiej detekcji błędów produkcyjnych i reagowania na nie, m.in. poprzez odpowiednio wczesne podejmowanie interwencji serwisowych lub przerywanie

pracy instalacji w chwili generowania błędów funkcjonalnych wytwarzanych produktów. Dodatkowo, metoda w naturalny sposób wykorzystuje odchylenia od referencyjnych wartości cech wzorców, jakimi są centroidy, stanowi więc uzupełnienie dla metody wykrywania wzorców technologicznych opartych o zmodyfikowany algorytm *k*-Means przedstawiony w rozdziale 4. Współczesne metody wykrywania anomalii w procesach produkcyjnych, bazujące na danych historycznych obejmują zwykle operacje na przebiegach czasowych, a jeśli oparte są o metody klasteryzacji, to dotyczą bardzo wąskiej grupy aplikacji, w szczególności wykrywania nieprawidłowości w pracy układów napędowych i badania stopnia degradacji ich elementów. Brak jest szeroko opisanych metod możliwych do zastosowania na ogólnie zdefiniowanych stanowiskach produkcyjnych, co stawia metodę na pozycji kandydata do zastosowania w różnych aplikacjach związanych z pracą w produkcji dyskretniej.

Pomimo, że eksperyment przygotowywany był dla symulacji wycieków przejawiających się zwiększonym zużyciem powietrza w układzie, zarejestrowano przy okazji co najmniej kilka cykli produkcyjnych w których wykryto nieprawidłowości wynikające z zablokowania mechanicznego produktu wewnątrz stacji, co wpłynęło na wydłużenie czasu pracy wybranych urządzeń wykonawczych. Błędy w pracy, zarówno te wynikające z nieprawidłowych wartości zużycia powietrza, jak i nieprawidłowych czasów pracy urządzeń wykonawczych, zostały prawidłowo wykryte i sklasyfikowane. Metoda bazuje na wykorzystaniu właściwości odchylenia standardowego poszczególnych cech wzorców technologicznych, co w przypadku nauki wzorców na niezaburzonych danych referencyjnych umożliwia uzyskanie dużej czułości na występujące anomalie. Chociaż metoda uruchomiona i zwalidowana została na danych przetworzonych w trybie post-procesowym, ze względu na niską złożoność obliczeniową, może być z powodzeniem stosowana do bieżącej detekcji anomalii, również na poziomie procedur sterujących PLC w trybie czasu rzeczywistego.

Ograniczeniem metody jest konieczność uprzedniej znajomości badanego wzorca technologicznego. W przypadku analizowania cyklu produkcyjnego odpowiadającego nieznanemu wzorcowi, metoda sklasyfikuje go jako anomalię. Z tego powodu analiza danych zebranych bezpośrednio podczas pojawiania się nowych wariantów produkcyjnych będzie nieskuteczna, stanowić jednak może wskazanie na potencjalną naukę nowego wzorca.

Dalsze wykorzystanie metody może być rozbudowane o analizę przyrostu wartości odstających: monitorowanie odchyleń poszczególnych cech wektorów V_{MCn} pozwolić może na określenie trendów ich zmian w czasie, a w przypadku wykrycia postępujących pogorszeń, wskazywać może na degradację wybranych fragmentów stanowiska produkcyjnego. Informacje te posłużyć mogą do bieżącego usuwania nieprawidłowości w działaniu stanowisk, co prowadzić może do efektywniejszego wykorzystania infrastruktury produkcyjnej oraz zwiększenia wydajności produkcji.

Prace badawcze skoncentrowane wokół zagadnień opisanych w tym rozdziale zostały przedstawione i opublikowane w czasopismach naukowych i podczas konferencji:

- Cupek, R., Ziebinski, A., Zonenberg, D., & Drewniak, M. (2018). Determination of the machine energy consumption profiles in the mass-customised manufacturing. *International Journal of Computer Integrated Manufacturing*, 31(6), 537-561.
- Cupek, R., Ziębiński, A., Drewniak, M., & Fojcik, M. (2018, March). Improving KPI based performance analysis in discrete, multi-variant production. In *Asian Conference on Intelligent Information and Database Systems* (pp. 661-673). Springer, Cham.
- Cupek, R., Ziębiński, A., Drewniak, M., & Fojcik, M. (2018, March). Estimation of the Number of Energy Consumption Profiles in the Case of Discreet Multi-variant Production. In *Asian Conference on Intelligent Information and Database Systems* (pp. 674-684). Springer, Cham.

6. ZAKOŃCZENIE

Niniejsza praca prezentuje wyniki badań autora prowadzone przy opracowywaniu metod analizy efektywności i jakości procesów produkcji dyskretnej, krótkoseryjnej i wielowariantowej z wykorzystaniem detekcji i klasyfikacji wzorców technologicznych do automatycznego wykrywania anomalii występujących podczas pracy stanowisk oraz nad możliwościami powiązania tych analiz z wiedzą technologiczną tak, aby wspierać zarządzanie i utrzymanie instalacji. Opracowane metody wykorzystują techniki z obszaru eksploracji danych, w szczególności zaproponowaną przez autora zmodyfikowaną metodę klasteryzacji *k*-Means, statystyki, modele informacyjne oparte o standard ANSI/ISA-95 i język Automation Markup Language oraz interfejs komunikacyjny OPC UA. W ocenie autora najważniejszym rezultatem opracowanych metod jest zdolność do adaptacyjnego i automatycznego wykrywania nowych wzorców technologicznych pojawiających się podczas występowania nowoprodukowanych wariantów oraz możliwość integracji danych procesowych i wyników prowadzonych na nich analiz z wiedzą dziedzinową dotyczącą wykonania instalacji produkcyjnej oraz wytwarzanych wariantów. Podejście to jest zgodne z filozofią nowoczesnych systemów produkcyjnych, mających wspierać dynamicznie zmieniającą się produkcję wielowariantową oraz umożliwiającą powiązanie różnych aspektów produkcji, od konstrukcji stanowisk, przez przepływ danych w fazie realizacji produkcji po utrzymanie infrastruktury. Opracowane metody, oprócz zapewnienia tej adaptacyjności poprzez automatyczne wykrywanie nowych wzorców technologicznych, wpisują się w referencyjną architekturę RAMI4.0 utworzoną dla nowoczesnych systemów produkcyjnych. Metody mogą być z powodzeniem stosowane dla każdej instalacji produkcyjnej pracującej w produkcji dyskretnej, która wyposażona jest w pomiar zużycia energii oraz dla której możliwy jest dostęp do pomiaru czasu aktywności poszczególnych urządzeń wykonawczych lub sygnałów aktywacji tych urządzeń. Dzięki specyfice zastosowanych algorytmów metody analityczne nie wymagają znajomości technologii, tj. podawania z góry informacji na temat produkowanych wariantów oraz fizycznej budowy stanowiska produkcyjnego.

Badania opisane w rozdziale 3. wskazują, że powiązanie danych procesowych i rezultatów z analizami efektywności i jakości realizowanych procesów z danymi technologicznymi stanowi otwarcie dla nowego typu przepływu danych, dzięki któremu cyfrowy bliźniak instalacji i prowadzonego procesu może być zasilony danymi z faktycznie prowadzonych czynności produkcyjnych. Uzupełnia to przepływ informacji

związany z utrzymaniem instalacji, umożliwiając nie tylko bezpośrednie wskazanie na technologię (urządzenia wykonawcze lub ich składowe) dla której wykrywane są anomalie, ale dodatkowo wiąże miejsce występowania anomalii z zasobami, które mogą zostać wykorzystane w procesie utrzymania lub poprawy technologii, m.in. plikami projektowymi czy dokumentami technicznymi. Wobec rezultatów otrzymanych w trakcie badań i opisanych w rozdziale 3. oraz poprzez umożliwienie powiązania danych o procesie z danymi o technologii w celu prowadzenia dalszych analiz opisanych w rozdziałach 4. i 5. dowiedziona została prawdziwość tezy głównej:

Zastosowanie modeli informacyjnych do powiązania danych procesowych z wiedzą technologiczną i wykorzystanie mechanizmów eksploracji danych pozwala na istotną poprawę efektywności pracy i utrzymania dostępności instalacji przemysłowych.

Badania opisane w rozdziale 4. opisują metodę automatycznego wykrywania wzorców technologicznych, zakładającą działanie na odpowiednio zagregowanych i przygotowanych do analizy wektorach danych opisujących cykle produkcyjne. Dzięki temu ilość danych poddawanych dalszej analizie ulega znaczącemu uproszczeniu, a każdy wykrywany wzorzec technologiczny opisany jest skończonym zestawem wartości wynikającym z zastosowanej technologii. Ponieważ do konstrukcji wektorów wykorzystane zostały podstawowe informacje o zużyciu energii i aktywności poszczególnych urządzeń wykonawczych, metoda jest uniwersalna i możliwa do zastosowania w różnych aplikacjach, niezależnie od technologii wykonania badanych instalacji. Wynika to z faktu ogólnej dostępności tych wielkości na poziomie jednostek kontrolnych procesu. Co ważne, agregacja danych realizowana jest w oparciu o proste mechanizmy wyznaczania wartości. Umożliwia to konstruowanie wektorów do analizy bezpośrednio po zakończonych cyklach produkcyjnych, a przez to dalsze wykorzystanie na zewnętrznych jednostkach obliczeniowych w trybie bieżącym, m.in. do adaptacyjnego wyznaczania nowych wzorców w chwili pojawienia się nowo produkowanego wariantu produkcyjnego. Zagadnienia opisane w rozdziale 4. dowiodły prawdziwości dwóch tez dodatkowych:

Agregacja danych wyliczanych w obrębie cykli produkcyjnych pozwala na zredukowanie złożoności obliczeniowej i skrócenie czasu realizacji algorytmów wyszukiwania wzorców w produkcji dyskretnej wielowariantowej.

oraz

Możliwe jest automatyczne wykrywanie i klasyfikacja profili produkcyjnych instalacji bez znajomości technologii a priori (tj. bez wiedzy na temat stosowanych urządzeń i specyfiki ich pracy, bez znajomości procesu produkcyjnego oraz bez wiedzy na temat aktualnie produkowanego typu produktu), jedynie z wykorzystaniem informacji o zużywanych energiach, czasach trwania i sygnałach binarnych definiujących aktywność urządzeń wykonawczych.

W rozdziale 5. opisana została metoda automatycznej klasyfikacji cykli produkcyjnych i automatycznego wykrywania anomalii w pracy urządzeń, wykorzystująca wartości wektorów opisujących wyznaczone wzorce technologiczne. Dopasowanie obserwacji bieżących do referencyjnych odbywa się za pomocą mechanizmów statystycznych, a wykrywanie nieprawidłowości badane jest w oparciu o naturalne właściwości centroidów wyznaczonych klastrów odpowiadających wzorcom. Ze względu na swoją szybkość metoda, podobnie jak wyznaczanie wartości zagregowanych, nadaje się do zastosowania na poziomie procedur sterujących kontrolerów procesu, dzięki czemu klasyfikacja i wykrywanie anomalii może być prowadzone w trybie czasu rzeczywistego. Otwiera to możliwości szybkiej detekcji błędów produkcyjnych i reagowania na nie, m.in. poprzez odpowiednio wczesne podejmowanie interwencji serwisowych lub przerywanie pracy instalacji w chwili generowania błędów funkcjonalnych wytwarzanych produktów. Rezultaty z prac badawczych opisanych w rozdziale 5. udowodniły prawdziwość trzeciej tezy dodatkowej:

Zaproponowane podejście do badania efektywności i jakości procesów produkcyjnych, bazujące na agregacji danych i automatycznym tworzeniu profili, umożliwia wykrywanie anomalii w pracy instalacji przemysłowych w produkcji dyskretnej.

Zaproponowane metody cechują się dużą prostotą w porównaniu do innych rozwiązań przedstawionych w rozdziałach 2.5., 2.6. oraz 2.7. Wynika to z faktu zastosowania prostych mechanizmów agregacji i przetwarzania analitycznego, a bezpośrednio przekłada się na szybkość realizowanych obliczeń i możliwość implementacji części z nich na poziomie kontrolerów procesu. Dodatkowo w żaden sposób nie wymagają zasilenia wiedzą

na temat technologii wykonania instalacji. Jednocześnie posiadają kilka ograniczeń. Trening nowych wzorców technologicznych prowadzony powinien być w oparciu o dane bez zaburzeń, w przeciwnym razie wyznaczone zostaną nadmiarowe lub nieprawidłowe wzorce. Ponieważ wzorce służą w dalszych krokach jako referencja do analizy bieżących cykli produkcyjnych, prowadzić to może do nieprawidłowej klasyfikacji i błędnej ich oceny, skutkującej błędnym wykrywaniem anomalii w pracy stanowiska produkcyjnego. Dodatkowym utrudnieniem może być modelowanie instalacji produkcyjnej z wykorzystaniem języka AutomationML lub materiałów produkcyjnych w oparciu o standard ANSI/ISA-95. Ponieważ nie istnieją gotowe rozwiązania wspierające szybkie tworzenie tych modeli oraz z uwagi na unikalność poszczególnych systemów produkcyjnych, ich budowanie jest każdorazowo indywidualnym procesem, mogącym cechować się dużą czasochłonnością i, w zależności od poziomu szczegółowości modelowanych zasobów, stopniem skomplikowania. Modele wymagają ponadto systemu produkcyjnego, który posiadać będzie możliwość ich odczytu, parsowania danych i wykorzystywania do zarządzania, co wprowadza dodatkowy element integracji.

Wprowadzenie zaproponowanych metod do środowisk produkcyjnych może usprawnić procesy produkcyjne poprzez skuteczniejsze utrzymanie instalacji produkcyjnych. Wczesne wykrywanie nieprawidłowości w pracy wskazywać może na występowanie anomalii, uruchamiać procedury utrzymaniowe i serwisowe oraz wspierać zarządzanie wiedzą o technologii, np. przez uzupełnianie dzienników zdarzeń dla konkretnych urządzeń produkcyjnych czy poprawę doboru odpowiednich nastaw i materiałów dla wybranych wariantów produkcyjnych. Dalszy rozwój przez analizę zmian i trendów w czasie otwiera możliwości wczesnego wykrywania postępujących degradacji urządzeń i prowadzić może do wydłużenia czasów bezawaryjności stanowisk produkcyjnych. Dodatkowo, transfer wiedzy wynikający z powiązania rezultatów działania mechanizmów analitycznych z wiedzą o technologii wspierać może procesy projektowania stanowisk czy kroków procesowych.

Opracowane metody i osiągnięte rezultaty potwierdzają prawdziwość tezy głównej rozprawy. Rezultaty z badań opisanych w rozdziale 6. wskazują na możliwość poprawy efektywności pracy i utrzymania dostępności instalacji przemysłowych poprzez zastosowanie modeli informacyjnych do powiązania danych procesowych z wiedzą technologiczną. Dodatkowo, w rozdziale 4. wykazano, że analiza danych procesowych

i skuteczne wykorzystanie wyników tej analizy możliwe jest bez znajomości wiedzy o technologii a priori, dzięki czemu zakres stosowalności ograniczony jest wymaganiami dotyczącymi specyfiki odpowiednich danych procesowych na wejściu algorytmów. Metody wykorzystują uproszczone zestawy danych zagregowanych, znacząco redukując wymaganą moc obliczeniową i ograniczając czas przetwarzania, co zapewnia możliwość implementacji części z nich nawet na poziomie kontrolerów procesu działających w reżimie czasu rzeczywistego. Zostało to przedstawione w rozdziałach 4. oraz 5. Finalnie, dzięki zastosowaniu wzorców technologicznych i uproszczonej formy danych do przetwarzania możliwe jest automatyczne wykrywanie anomalii produkcyjnych, co opisane zostało w rozdziale 5.

Współcześnie stosowane techniki eksploracji danych opartych o klasteryzację cieszą się dużą popularnością w różnych dziedzinach, w tym coraz częściej w przemyśle. Stosowanie poszczególnych algorytmów i ich modyfikacje prowadzone są zwykle pod kątem specyfiki aplikacji i niejednokrotnie są przez to ograniczone w stosowaniu do pojedynczej lub mocno ograniczonej grupy rozwiązań. Metody opisane w niniejszej rozprawie generalizują podejście do poszukiwania odpowiedzi procesy na mnogość i różnorodność produkowanych wariantów, a przez to zapewniają szerszy zakres stosowalności. Dodatkowo, w dobie rozwoju systemów wspierających produkcję elastyczną i krótkoseryjną, zaproponowane wykorzystanie modeli informacyjnych zgodnych z architekturą RAMI4.0 prezentuje praktyczną stronę ich aplikowalności, tj. dla kompleksowego rozwiązania działającego na różnych poziomach struktury produkcyjnej, wykorzystującego różne aspekty produkcji od wytwarzania po utrzymanie oraz prezentującego przepływ danych, informacji i wiedzy dla różnych potrzeb i specyfik. Rozwiązań tego typu w dziedzinie informatyki jest niewiele, głównie ze względu na młodość trendów informatycznych panujących w branży produkcyjnej oraz ze względu na złożoność i wspomnianą kompleksowość łączącą cechy przetwarzania analitycznego, zarządzania produkcją, zarządzania wiedzą i wspierania procesów utrzymania produkcji.

Zaproponowane rozwiązanie, choć sprawdzone i badane było na rzeczywistej instalacji produkcyjnej, w różnych środowiskach może oczywiście wymagać dostosowania lub wprowadzenia modyfikacji. Każda instalacja i każda branża produkcyjna posiada własne, unikatowe cechy, mogące wpływać na jakość, specyfikę i możliwość wykorzystania danych procesowych do analiz. Głównym obiektem badań było stanowisko produkcyjne realizujące produkcję w trybie sekwencyjnym, w którym poszczególne kroki prowadzone

były jeden po drugim. Dla przypadków instalacji sterowanych w sposób kombinacyjny mnogość kombinacji sygnałów może wpływać na aplikowalność przebadanych metod lub wręcz uniemożliwiać ich zastosowanie. Dalszy ich rozwój wymagać więc może badania dla złożonych instalacji produkcyjnych oraz modyfikacji na poziomie mechanizmów automatycznej detekcji wzorców i automatycznej detekcji anomalii. Otwiera to jednocześnie nowe kierunki badań i perspektywy dalszego rozwoju, co w połączeniu z dynamicznym rozwojem i stosowaniem nowoczesnych modeli informacyjnych prowadzić może do coraz pełniejszego wsparcia branży produkcyjnej narzędziami informatycznymi.

7. BIBLIOGRAFIA

- [1] Bagnall, Anthony; Lines, Jason; Bostrom, Aaron; Large, James; Keogh, Eamonn (2017): The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. In *Data mining and knowledge discovery* 31 (3), pp. 606–660.
- [2] Miljković, Dubravko (2011): Fault detection methods: A literature survey. In : 2011 Proceedings of the 34th international convention MIPRO: IEEE, pp. 750–755.
- [3] Park, You-Jin; Fan, Shu-Kai S.; Hsu, Chia-Yu (2020): A review on fault detection and process diagnostics in industrial processes. In *Processes* 8 (9), p. 1123.
- [4] Isermann, Rolf (1984): Process fault detection based on modeling and estimation methods - a survey. In *automatica* 20 (4), pp. 387–404.
- [5] Merayo, D.; Rodriguez-Prieto, A.; am Camacho (2019): Comparative analysis of artificial intelligence techniques for material selection applied to manufacturing in Industry 4.0. In *Procedia manufacturing* 41, pp. 42–49.
- [6] Chien, Chen-Fu; Wang, Wen-Chih; Cheng, Jen-Chieh (2007): Data mining for yield enhancement in semiconductor manufacturing and an empirical study. In *Expert Systems with Applications* 33 (1), pp. 192–198.
- [7] Luhn, Hans Peter (1958): A business intelligence system. In *IBM Journal of research and development* 2 (4), pp. 314–319.
- [8] Harrington, Joseph (1979): Computer integrated manufacturing: RE Krieger Publishing Company.

- [9] Saenz de Ugarte, B.; Artiba, A.; Pellerin, R. (2009): Manufacturing execution system - a literature review. In *Production planning and control* 20 (6), pp. 525–539.
- [10] Arica, Emrah; Powell, D. J. (2017): Status and future of manufacturing execution systems. In : 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM): IEEE, pp. 2000–2004.
- [11] STANDARD, DRAFT (2008): Enterprise-Control System Integration Part 1: Models and Terminology.
- [12] Zhou, Zude; Xiao, Zheng; Liu, Quan; Ai, Qingsong (2013): An analytical approach to customer requirement information processing. In *Enterprise Information Systems* 7 (4), pp. 543–557.
- [13] Tilbury, Dawn M. (2019): Cyber-physical manufacturing systems. In *Annual Review of Control, Robotics, and Autonomous Systems* 2, pp. 427–443.
- [14] Tseng, Mitchell M.; Jiao, Jianxin (2001): Mass customization. In *Handbook of industrial engineering* 3, pp. 684–709.
- [15] Harnly, Aaron; Nenkova, Ani; Passonneau, Rebecca; Rambow, Owen (2005): Automation of summary evaluation by the pyramid method. In : Recent Advances in Natural Language Processing (RANLP), pp. 226–232.
- [16] Schnicke, Frank; Kuhn, Thomas; Antonino, Pablo Oliveira (2020): Enabling industry 4.0 service-oriented architecture through digital twins. In : European Conference on Software Architecture: Springer, pp. 490–503.
- [17] Fraile, Francisco; Sanchis, Raquel; Poler, Raul; Ortiz, Angel (2019): Reference models for digital manufacturing platforms. In *Applied Sciences* 9 (20), p. 4433.
- [18] Song, Hang; Jiang, Yuncheng (2020): The value chain of Industrial IoT and its reference framework for digitalization. In *arXiv preprint arXiv:2009.13039*.
- [19] Ahmadi, Ahmadzai; Cherifi, Chantal; Cheutet, Vincent; Ouzrout, Yacine (2020): Recent advancements in smart manufacturing technology for modern industrial revolution: A survey. In *Journal of Engineering and Information Science Studies*.
- [20] Fraile, Francisco; Sanchis, Raquel; Poler, Raul; Ortiz, Angel (2019): Reference models for digital manufacturing platforms. In *Applied Sciences* 9 (20), p. 4433.
- [21] Mobley, R. Keith (2002): An introduction to predictive maintenance: Elsevier.

- [22] Williams, Theodore J. (1990): A reference model for computer integrated manufacturing from the viewpoint of industrial automation. In *IFAC Proceedings Volumes* 23 (8), pp. 281–291.
- [23] Williams, T. J. (1992): The Purdue Enterprise Reference Architecture, Instrument Society of America. In *Research Triangle Park*.
- [24] Cupek, Rafal; Ziebinski, Adam; Huczala, Lukasz; Erdogan, Huseyin (2016): Agent-based manufacturing execution systems for short-series production scheduling. In *Computers in Industry* 82, pp. 245–258.
- [25] Neumann, Peter (2007): Communication in industrial automation - What is going on? In *Control Engineering Practice* 15 (11), pp. 1332–1347.
- [26] Mahnke, Wolfgang; Leitner, Stefan-Helmut; Damm, Matthias (2009): OPC unified architecture: Springer Science & Business Media.
- [27] Lehnhoff, Sebastian; Rohjans, Sebastian; Uslar, Mathias; Mahnke, Wolfgang (2012): OPC unified architecture: A service-oriented architecture for smart grids. In : 2012 First International Workshop on Software Engineering Challenges for the Smart Grid (SE-SmartGrids): IEEE, pp. 1–7.
- [28] Leitner, Stefan-Helmut; Mahnke, Wolfgang (2006): OPC UA: service-oriented architecture for industrial applications. In *ABB Corporate Research Center* 48 (61-66), p. 22.
- [29] Lange, Jurgen; Iwanitz, Frank; Burke, Thomas J. (2010): OPC: von data access bis unified architecture: VDE.
- [30] Henssen, Robert; Schleipen, Miriam (2014): Interoperability between OPC UA and AutomationML. In *Procedia Cirp* 25, pp. 297–304.
- [31] Ono, Toshio; Ali, Shahzad; Hunkar, Paul; Brandl, Dennis (2013): OPC Unified Architecture for ISA-95 Common Object Model. In *Companion Specification Release 1*.
- [32] Cupek, Rafal; Ziebinski, Adam; Franek, Maciej (2013): FPGA based OPC UA embedded industrial data server implementation. In *Journal of Circuits, Systems and Computers* 22 (08), p. 1350070.
- [33] Larose, T. D. (2006): Odkrywanie wiedzy z danych, Wyd. In *Naukowe PWN, Warszawa*.

- [34] Aggarwal, Charu C. (2015): *Data mining: the textbook*: Springer (1).
- [35] Lavine, Barry K.; Mirjankar, Nikhil (2000): Clustering and classification of analytical data. In *Encyclopedia of analytical chemistry* 11.
- [36] Kingsford, Carl; Salzberg, Steven L. (2008): What are decision trees? In *Nature biotechnology* 26 (9), pp. 1011–1013.
- [37] Sharma, Meenakshi (2014): Data mining: A literature survey. In *International Journal of Emerging Research in Management & Technology* 3 (2).
- [38] Ge, Zhiqiang; Song, Zhihuan; Ding, Steven X.; Huang, Biao (2017): Data mining and analytics in the process industry: The role of machine learning. In *Ieee Access* 5, pp. 20590–20616.
- [39] Hand, David J. (2007): Principles of data mining. In *Drug safety* 30 (7), pp. 621–622.
- [40] Olson, David L. (2007): Data mining in business services. In *Service Business* 1 (3), pp. 181–193.
- [41] Tryon, Robert C. (1939): Cluster analysis: correlation profile and orthometric analysis for the isolation of unities in mind and personality. In *Ann Arbor: Edward Brothers*.
- [42] Grover, Nidhi (2014): A study of various fuzzy clustering algorithms. In *International Journal of Engineering Research* 3 (3), pp. 177–181.
- [43] Kameshwaran, K.; Malarvizhi, K. (2014): Survey on clustering techniques in data mining. In *International Journal of Computer Science and Information Technologies* 5 (2), pp. 2272–2276.
- [44] Rokach, Lior; Maimon, Oded (2005): Clustering methods. In : *Data mining and knowledge discovery handbook*: Springer, pp. 321–352.
- [45] Madhulatha, T. Soni (2012): An overview on clustering methods. In *arXiv preprint arXiv:1205.1117*.
- [46] Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (1967). 14th ed.: Oakland, CA, USA.

- [47] Kaufman, L.; Rousseeuw, P. J.; Dodge, Y. (1987): Clustering by means of medoids in statistical data analysis based on the norm and related methods. In *L1 Norm, ~ orth-Holland, Amsterdam*.
- [48] Ward Jr, Joe H. (1963): Hierarchical grouping to optimize an objective function. In *Journal of the American statistical association* 58 (301), pp. 236–244.
- [49] Dunn, Joseph C. (1973): A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. In *0022-0280*.
- [50] Fu, Limin; Medico, Enzo (2007): FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. In *BMC bioinformatics* 8 (1), pp. 1–15.
- [51] Ester, Martin; Kriegel, Hans-Peter; Sander, J. J.; Xu, Xiaowei (1996): A density-based algorithm for discovering clusters in large spatial databases with noise. In : *kdd*, vol. 96. 34th ed., pp. 226–231.
- [52] Ankerst, Mihael; Breunig, Markus M.; Kriegel, Hans-Peter; Sander, J. J. (1999): OPTICS: Ordering points to identify the clustering structure. In *ACM Sigmod record* 28 (2), pp. 49–60.
- [53] Fraley, Chris; Raftery, Adrian E. (1998): How many clusters? Which clustering method? Answers via model-based cluster analysis. In *The computer journal* 41 (8), pp. 578–588.
- [54] Dempster, Arthur P.; Laird, Nan M.; Rubin, Donald B. (1977): Maximum likelihood from incomplete data via the EM algorithm. In *Journal of the Royal Statistical Society: Series B (Methodological)* 39 (1), pp. 1–22.
- [55] MacQueen, James (1967): Some methods for classification and analysis of multivariate observations. In : *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1. 14th ed.: Oakland, CA, USA, pp. 281–297.
- [56] Mashor, Mohd Yusoff (2000): Hybrid training algorithm for RBF network. In *International Journal of the computer, the Internet and Management* 8 (2), pp. 50–65.
- [57] Bhatia, Sanjiv K. (2004): Adaptive K-Means Clustering. In : *FLAIRS conference*, pp. 695–699.

- [58] Schubert, Erich; Rousseeuw, Peter J. (2019): Faster k -medoids clustering: improving the PAM, CLARA, and CLARANS algorithms. In : International conference on similarity search and applications: Springer, pp. 171–187.
- [59] Tripathy, B. K.; Ghosh, Adhir; Panda, G. K. (2012): Kernel based K -means clustering using rough set. In : 2012 International Conference on Computer Communication and Informatics: IEEE, pp. 1–5.
- [60] Dexi, Kong; Rui, Kong (2013): A fast and effective kernel-based k -means clustering algorithm. In : 2013 Third International Conference on Intelligent System Design and Engineering Applications: IEEE, pp. 58–61.
- [61] Liu, Bowen; Zhang, Ting; Li, Yujian; Liu, Zhaoying; Zhang, Zhilin (2021): Kernel probabilistic k -means clustering. In *Sensors* 21 (5), p. 1892.
- [62] Jung, Yoon Mo; Whang, Joyce Jiyong; Yun, Sangwoon (2020): Sparse probabilistic K -means. In *Applied Mathematics and Computation* 382, p. 125328.
- [63] Lingras, Pawan; Yan, Rui; West, Chad (2003): Comparison of conventional and rough k -means clustering. In : International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing: Springer, pp. 130–137.
- [64] Lingras, Pawan (2009): Evolutionary rough K -means clustering. In : International Conference on Rough Sets and Knowledge Technology: Springer, pp. 68–75.
- [65] Bholowalia, Purnima; Kumar, Arvind (2014): EBK-means: A clustering technique based on elbow method and k -means in WSN. In *International Journal of Computer Applications* 105 (9).
- [66] Kaufman, Leonard; Rousseeuw, Peter J. (2009): Finding groups in data: an introduction to cluster analysis: John Wiley & Sons.
- [67] Akaike, Hirotugu (1974): A new look at the statistical model identification. In *IEEE transactions on automatic control* 19 (6), pp. 716–723.
- [68] Schwarz, Gideon (1978): Estimating the dimension of a model. In *The annals of statistics*, pp. 461–464.
- [69] Smyth, Padhraic (1996): Clustering Using Monte Carlo Cross-Validation. In : *Kdd*, vol. 1, pp. 26–133.

- [70] Ben-David, S.; Luxburg, U. von; Pal, D. (2006): A sober look at stability of clustering. In : Proceedings of the Annual Conference on Computational Learning Theory, vol. 10.
- [71] Ben-Hur, Asa; Elisseeff, Andre; Guyon, Isabelle (2001): A stability based method for discovering structure in clustered data. In : Biocomputing 2002: World Scientific, pp. 6–17.
- [72] Keogh, Eamonn; Chu, Selina; Hart, David; Pazzani, Michael (2004): Segmenting time series: A survey and novel approach. In : Data mining in time series databases: World Scientific, pp. 1–21.
- [73] Liu, Xiaoyan; Lin, Zhenjiang; Wang, Huaiqing (2008): Novel online methods for time series segmentation. In *IEEE Transactions on knowledge and data engineering* 20 (12), pp. 1616–1626.
- [74] Chung, Fu-lai; Fu, Tak-chung; Luk, Robert; Ng, Vincent (2002): Evolutionary time series segmentation for stock data mining. In : 2002 IEEE International Conference on Data Mining, 2002. Proceedings: IEEE, pp. 83–90.
- [75] Wang, Zhiguang; Yan, Weizhong; Oates, Tim (2017): Time series classification from scratch with deep neural networks: A strong baseline. In : 2017 International joint conference on neural networks (IJCNN): IEEE, pp. 1578–1585.
- [76] Susto, Gian Antonio; Cenedese, Angelo; Terzi, Matteo (2018): Time-series classification methods: Review and applications to power systems data. In *Big data application in power systems*, pp. 179–220.
- [77] Polge, Julien; Robert, J |[®]r |[®]my; Le Traon, Yves (2020): A case driven study of the use of time series classification for flexibility in industry 4.0. In *Sensors* 20 (24), p. 7273.
- [78] Sihag, Nitesh; Sangwan, Kuldip Singh (2020): Development of an Electric-Load Intelligence System for Component Level Disaggregation to Improve Energy Efficiency of Machine Tools. In : Enhancing Future Skills and Entrepreneurship: Springer, Cham, pp. 117–129.
- [79] Sihag, Nitesh; Sangwan, Kuldip Singh; Pundir, Siddhant (2018): Development of a structured algorithm to identify the status of a machine tool to improve energy and time efficiencies. In *Procedia Cirp* 69, pp. 294–299.

- [80] Seevers, Jan-Peter; Jurczyk, Kristina; Meschede, Henning; Hesselbach, Jens; Sutherland, John W. (2020): Automatic detection of manufacturing equipment cycles using time series. In *Journal of Computing and Information Science in Engineering* 20 (3).
- [81] Labbus, Ingo; Teiwes, Hanno; Filz, Marc-Andr |®; Herrmann, Christoph; Gonter, Mark; R |Âssinger, Markus; Thiede, Sebastian (2019): Automated statistical evaluation of energy data in the automotive production. In *Procedia Cirp* 81, pp. 1154–1159.
- [82] Carvajal Soto, J. A.; Tavakolizadeh, F.; Gyulai, D |ívid (2019): An online machine learning framework for early detection of product failures in an Industry 4.0 context. In *International Journal of Computer Integrated Manufacturing* 32 (4-5), pp. 452–465.
- [83] Pang, Chee Khiang; Le, Cao Vinh; Gan, Oon Peen; Chee, Xiang Min; Zhang, Dan Hong; Luo, Ming et al. (2011): Intelligent energy audit and machine management for energy-efficient manufacturing. In : 2011 IEEE 5th International Conference on Cybernetics and Intelligent Systems (CIS): IEEE, pp. 142–147.
- [84] Seevers, Jan-Peter; Johst, J.; Wei |f, Tim; Meschede, Henning; Hesselbach, Jens (2019): Automatic time series segmentation as the basis for unsupervised, non-intrusive load monitoring of machine tools. In *Procedia Cirp* 81, pp. 695–700.
- [85] Teiwes, Hanno; Blume, Stefan; Herrmann, Christoph; R |Âssinger, Markus; Thiede, Sebastian (2018): Energy load profile analysis on machine level. In *Procedia Cirp* 69, pp. 271–276.
- [86] Batako, Andre D. L.; Goh, Siew Y. (2014): Workpiece roundness profile in the frequency domain: an application in cylindrical plunge grinding. In *The International Journal of Advanced Manufacturing Technology* 72 (1), pp. 277–288.
- [87] Viera, M. AulestiaA; Alexandre, F. A.; Aguiar, P. R.; Silva, R. B.; Bianchi, E. C. (2018): Correlation between surface roughness and AE signals in ceramic grinding based on spectral analysis. In : MATEC Web of Conferences, vol. 249: EDP Sciences, p. 3003.
- [88] Rapur, J. S.; Tiwari, Rajiv (2020): An intelligent and robust fault diagnosis system for identification of centrifugal pump defects in frequency domain using corrupted vibration and current signatures. In : *Advances in Rotor Dynamics, Control, and Structural Health Monitoring*: Springer, pp. 407–426.

- [89] Ranawat, Nagendra Singh; Kankar, Pavan Kumar; Miglani, Ankur (2021): Fault Diagnosis in Centrifugal Pump using Support Vector Machine and Artificial Neural Network. In *Journal of Engg. Research EMSME Special Issue pp 99*, p. 111.
- [90] Shevchik, Sergey A.; Kenel, Christoph; Leinenbach, Christian; Wasmer, Kilian (2018): Acoustic emission for in situ quality monitoring in additive manufacturing using spectral convolutional neural networks. In *Additive Manufacturing 21*, pp. 598–604.
- [91] Petruschke, Lars; Burkhardt, Max; Kohne, Thomas; Schraml, Philipp; Helfert, Mark; Abele, Eberhard (2020): Method to identify energy efficiency potentials of metal cutting machine tools in industry. In *Procedia Cirp 90*, pp. 522–527.
- [92] Bornschlegl, Martin; Bregulla, Markus; Franke, J |Ârg (2016): Methods-Energy MeasurementÔÇôAn approach for sustainable energy planning of manufacturing technologies. In *Journal of Cleaner Production 135*, pp. 644–656.
- [93] Al-Sultan, Khaled S. (1997): A hard clustering approach to the part family formation problem. In *Production planning and control 8 (3)*, pp. 231–236.
- [94] Strasser, Sonja; Peirleitner, Andreas (2017): Reducing Variant Diversity by Clustering. In : *Proceedings of the 6th International Conference on Data Science, Technology and Applications*, pp. 141–148.
- [95] Gupta, Ashutosh; Jain, P. K.; Kumar, Dinesh (2014): A novel approach for part family formation for reconfiguration manufacturing system. In *Opsearch 51 (1)*, pp. 76–97.
- [96] Yiakopoulos, Christos; Gryllias, Konstantinos; Antoniadis, Ioannis (2009): Rolling element bearing fault classification using *K*-means frequency domain based clustering. In : *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 48982, pp. 127–134.
- [97] Amruthnath, Nagdev; Gupta, Tarun (2018): A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In : *2018 5th international conference on industrial engineering and applications (ICIEA): IEEE*, pp. 355–361.
- [98] Wu, Haixi; Yu, Zhonghua; Wang, Yan (2016): A new approach for online monitoring of additive manufacturing based on acoustic emission. In : *International*

Manufacturing Science and Engineering Conference, vol. 49910: American Society of Mechanical Engineers, V003T08A013.

- [99] Abd Majid, Nazatul Aini; Young, Brent R.; Taylor, Mark P.; Chen, John J. J. (2012): *K*-means clustering pre-analysis for fault diagnosis in an aluminium smelting process. In : 2012 4th Conference on Data Mining and Optimization (DMO): IEEE, pp. 43–46.
- [100] Karaca, Zeynep (2018): The cluster analysis in the manufacturing industry with *k*-means method: An application for Turkey. In *Eurasian Journal of Economics and Finance* 6 (3), pp. 1–12.
- [101] Hasanah, Uswatun; Humaidi, Bahrul (2020): Analysis of Data Mining Using *K*-Means Clustering Algorithm for Product Grouping. In *International Journal of Informatics and Information Systems* 3 (1), pp. 12–22.
- [102] Mardiana, Yessi; Kalsum, Toibah Umi (2022): Application Of Data Mining To Determine The Amount Of Production Using The *K*-Means Method (Case Study At Pt. Coca Cola Distribution Indonesia Bengkulu). In *GATOTKACA Journal (Teknik Sipil, Informatika, Mesin dan Arsitektur)* 3 (1), pp. 1–10.
- [103] Abbas, Aries; Prayitno, Pungkas; Nurkim, N.; Prumanto, Denny; Dewadi, Fathan Mubina; Hidayati, Nur; Windarto, Agus Perdana (2021): Implementation of clustering unsupervised learning using *K*-Means mapping techniques. In : IOP Conference Series: Materials Science and Engineering, vol. 1088. 1st ed.: IOP Publishing, p. 12004.
- [104] Chadha, Gavneet Singh; Krishnamoorthy, Monica; Schwung, Andreas (2019): Time series based fault detection in industrial processes using convolutional neural networks. In : IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society, vol. 1: IEEE, pp. 173–178.
- [105] Hsu, Chia-Yu; Liu, Wei-Chen (2021): Multiple time-series convolutional neural network for fault detection and diagnosis and empirical study in semiconductor manufacturing. In *Journal of Intelligent Manufacturing* 32 (3), pp. 823–836.
- [106] Sanchez-Fernandez, Alvar; Baldan, Francisco Javier; Sainz-Palmero, Gregorio Ismael; Benitez, Jose Manuel; Fuente, M. J. (2018): Fault detection based on time series modeling and multivariate statistical process control. In *Chemometrics and Intelligent Laboratory Systems* 182, pp. 57–69.

- [107] Sikora, Marek; Szczyrba, Krzysztof; Wróbel, Łukasz; Michalak, Marcin: Monitorowanie i utrzymanie suwnicy bramowej na podstawie bezprzewodowego systemu pomiaru i analizy poziomu drgań.
- [108] Le Nguyen, Cuong (2016): Adaptive Empirical Mode Decomposition for Bearing Fault Detection. In *Strojniski Vestnik/Journal of Mechanical Engineering* 62 (5).
- [109] Janczewski, Ł. (2011): Vibro-acoustic failure recognition on combustion engines at the end of assembly lines. In *Silniki Spalinowe* 50.
- [110] Dolenc, Bostjan; Boskoski, Pavle; Juricič, Dani (2016): Distributed bearing fault diagnosis based on vibration analysis. In *Mechanical Systems and Signal Processing* 66, pp. 521–532.
- [111] Patil, S. S.; Gaikwad, J. A. (2013): Vibration analysis of electrical rotating machines using FFT: A method of predictive maintenance. In : 2013 fourth international conference on computing, communications and networking technologies (ICCCNT): IEEE, pp. 1–6.
- [112] Aherwar, Amit (2012): An investigation on gearbox fault detection using vibration analysis techniques: A review. In *Australian Journal of Mechanical Engineering* 10 (2), pp. 169–183.
- [113] Puggini, Luca; Doyle, John; McLoone, Sean (2015): Fault detection using random forest similarity distance. In *IFAC-PapersOnLine* 48 (21), pp. 583–588.
- [114] Yang, Bo-Suk; Di, Xiao; Han, Tian (2008): Random forests classifier for machine fault diagnosis. In *Journal of mechanical science and technology* 22 (9), pp. 1716–1725.
- [115] Yaman, Orhan (2021): An automated faults classification method based on binary pattern and neighborhood component analysis using induction motor. In *Measurement* 168, p. 108323.
- [116] Wasnik, Prerana P.; Phadkule, N. J.; Thakur, K. D. (2019): Fault detection and classification based on semi-supervised machine learning using KNN. In : 2019 International Conference on Innovative Trends and Advances in Engineering and Technology (ICITAET): IEEE, pp. 79–83.

- [117] Huang, Yang; Chen, Chiun-Hsun; Huang, Chi-Jui (2019): Motor fault detection and feature extraction using RNN-based variational autoencoder. In *Ieee Access* 7, pp. 139086–139096.
- [118] Fan, Shu-Kai S.; Lin, Shou-Chih; Tsai, Pei-Fang (2016): Wafer fault detection and key step identification for semiconductor manufacturing using principal component analysis, AdaBoost and decision tree. In *Journal of Industrial and Production Engineering* 33 (3), pp. 151–168.
- [119] Han, Xianyao; Tian, Shengwei; Romagnoli, Jose A.; Li, Hui; Sun, Wei (2018): PCA-SDG based process monitoring and fault diagnosis: application to an industrial pyrolysis furnace. In *IFAC-PapersOnLine* 51 (18), pp. 482–487.
- [120] Ma, Liang; Dong, Jie; Peng, Kaixiang (2021): A Novel Multilabel Classification Framework for Coupling Faults in Hot Rolling Processes. In *IEEE Transactions on Control Systems Technology*.
- [121] Jiang, Mon-Fong; Tseng, Shian-Shyong; Su, Chih-Ming (2001): Two-phase clustering process for outliers detection. In *Pattern recognition letters* 22 (6-7), pp. 691–700.
- [122] Wei, Zexian; Wang, Yanxue; He, Shuilong; Bao, Jiading (2017): A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection. In *Knowledge-Based Systems* 116, pp. 1–12.
- [123] Yuwono, Mitchell; Qin, Yong; Zhou, Jing; Guo, Ying; Celler, Branko G.; Su, Steven W. (2016): Automatic bearing fault diagnosis using particle swarm clustering and Hidden Markov Model. In *Engineering Applications of Artificial Intelligence* 47, pp. 88–100.
- [124] Li, Guannan; Hu, Yunpeng (2018): Improved sensor fault detection, diagnosis and estimation for screw chillers using density-based clustering and principal component analysis. In *Energy and Buildings* 173, pp. 502–515.
- [125] Li, Xiang; Li, Xu; Ma, Hui (2020): Deep representation clustering-based fault diagnosis method with unsupervised data applied to rotating machinery. In *Mechanical Systems and Signal Processing* 143, p. 106825.
- [126] Cupek, Rafal; Folkert, Kamil; Huczala, Lukasz; Zonenberg, Dariusz; Tomczyk, Jakub (2013): End-point device compressed air consumption analysis by reverse

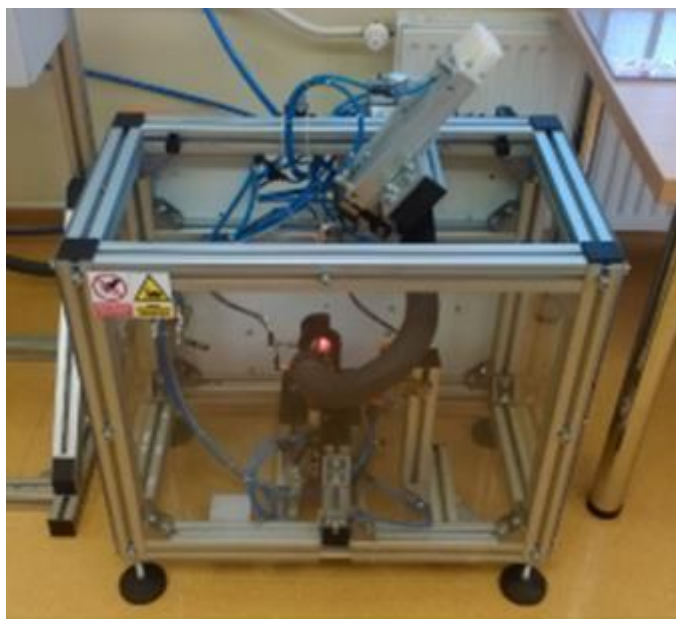
engineering algorithm. In : IECON 2013-39th Annual Conference of the IEEE Industrial Electronics Society: IEEE, pp. 7519–7524.

[127] Jaccard, James; Becker, Michael A. (2021): Statistics for the behavioral sciences: Cengage Learning.

ZAŁĄCZNIK 1 - OPIS STANOWISKA PRODUKCYJNEGO WYKORZYSTYWANEGO W TRAKCIE BADAŃ

Stanowisko DC40

Stanowisko testowe DC40 jest stanowiskiem eksperymentalnym znajdującym się w siedzibie firmy Aiut Sp. z o.o. w Gliwicach i powstało w celu badania efektywności energetycznej urządzeń bazujących na wykorzystaniu sprężonego powietrza. Urządzenie wykorzystywane jest do transportu detali w obiegu zamkniętym i składa się z rzeczywistych komponentów wykorzystywanych w rozwiązaniach automatyki przemysłowej. Podstawowymi elementami wykonawczymi w układzie są cztery siłowniki (dwa do transportu dwukierunkowego i dwa do dozowania przepływu detalu), elementu ssącego oraz dmuchawy. Stacja przygotowania powietrza doposażona została w przepływomierz badający chwilowy przepływ powietrza oraz w dodatkowe zawory umożliwiające odprowadzanie powietrza poza układ, umożliwiając symulację wycieków z przewodów pneumatycznych. Widok ogólny stanowiska został przedstawiony na rysunku 1.



Rysunek 1. – widok ogólny stanowiska DC40

Sekwencja transportowa i obsługa procesu realizowana jest przez sterownik PLC Siemens S7-300, obsługujący wyspy sygnałowe ET200L, ET200S oraz ET200X, do których podłączone są urządzenia wykonawcze oraz czujniki obecności detalu w układzie i pozycji poszczególnych siłowników. Stanowisko dodatkowo wyposażone zostało

w panel operatorski Siemens OP177b oraz zestaw przycisków, selektorów i lamp umiejscowionych na froncie szafy sterowniczej, tworzące interfejs HMI dla operatora.

W celu rejestracji danych procesowych na stacji opracowany i uruchomiony został system OPC UA. Sterownik PLC cyklicznie transmituje wybrane dane procesowe za pomocą protokołu TCP/IP, które następnie odbierane są przez serwer OPC UA, a który propaguje te dane do klientów końcowych. Rekord danych wykorzystywanych w badaniach autora wraz z opisem poszczególnych pól przedstawiony został w tabeli 1.

Tabela 1. – rekord danych rejestrowanych podczas pracy stanowiska DC40

Stempel czasowy PLC	Chwilowe zużycie powietrza	Całkowite (przyrostowe) zużycie powietrza	Stan wyjść cyfrowych PLC
Milisekunda	Litry/minutę	Litry	INT, dziesiętnie

Urządzenia wykonawcze monitorowane są w trakcie pracy, a ich aktywność definiowana jest za pomocą logicznych wartości pogrupowanych w odpowiednich bajtach odpowiadających sygnałom na wejściach i wyjściach sterownika PLC. Opis słowa wyjściowego zawierającego sterowanie urządzeniami pneumatycznymi DC40 zebrany został w tabeli 2.

Tabela 2. – Opis słowa wyjściowego PLC z urządzeniami wykonawczymi pneumatyki dla stanowiska DC40

Adres	Urządzenie
Q20.0	Podnośnik w górę
Q20.1	Podnośnik w dół
Q20.2	Wypychacz w przód
Q20.3	Wypychacz w tył
Q20.4	Pre-stop w dół
Q20.5	Stop w dół
Q20.6	Zassanie
Q20.7	Wydmuch
Q21.0	Zawór dodatkowy 1
Q21.1	Zawór dodatkowy 2
Q21.2	Zawór dodatkowy 3
Q21.3	Zawór dodatkowy 4
Q21.4	Wyjście rezerwowe 1
Q21.5	Wyjście rezerwowe 2
Q21.6	Wyjście rezerwowe 3
Q21.7	Wyjście rezerwowe 4

Stanowisko pracować może w różnych wariantach produkcyjnych, w których wykorzystywane są różne zestawy urządzeń wykonawczych, które dodatkowo różnić się mogą czasami ich wykorzystania. Istotnym jest również fakt, że niektóre urządzenia mogą pracować jednocześnie, co wynika nie tylko z realizacji konkretnych wariantów, ale również z możliwości obsługi wielu detali produkcyjnych jednocześnie.