

Recenzja spełnia wymogi formalne



Politechnika Warszawska
Wydział Transportu

Przewodniczący Rady Dyscypliny
Inżynieria Lądowa, Geodezja i Transport
Politechniki Śląskiej

prof. dr hab./inż. Piotr Folęga

Warszawa, dnia 31.03.2025

dr hab. inż. Andrzej Czerepicki, prof. PW
Zakład Systemów Informatycznych i Mechatronicznych w Transporcie

RECENZJA

Rozprawy doktorskiej mgr inż. Artura Budzyńskiego
pt. „Prognozowanie cen usług drogowego transportu towarowego z wykorzystaniem
uczenia maszynowego” (Forecasting prices for road freight transport services using
machine learning)

Podstawą wykonania recenzji jest Uchwała nr 84/2025 Rady Dyscypliny Inżynieria Lądowa,
Geodezja i Transport Politechniki Śląskiej z dnia 20 lutego 2025 r. oraz pismo
Przewodniczącego Rady dr hab. inż. Piotra Folęgi, profesora Politechniki Śląskiej z dnia 20
lutego 2025 r. (numer sprawy RDILGT.512.32.2025).

Dokumentację merytoryczną do opracowania recenzji stanowi wydrukowany egzemplarz
rozprawy mgr inż. Artura Budzyńskiego pt. „Forecasting prices for road freight transport
services using machine learning” napisany w języku angielskim, ze streszczeniem w języku
polskim.

Promotorem rozprawy jest prof. dr hab. inż. Aleksander Śladowski.

1. Wprowadzenie

Przedstawiona do oceny rozprawa doktorska dotyczy prognozowania cen na usługi
przewozów drogą lądową. Transport drogowy towarów stanowi jeden z ważniejszych
segmentów rynku przewozowego. Charakteryzuje się on między innymi: dostępnością
cenową, łatwością dostosowania się do potrzeb konkretnych kontrahentów, zdolnością do
przewiezienia ładunków o różnych wymiarach, wadze i konsystencji, względną szybkością
dostaw oraz zasięgiem obejmującym zarówno krótkie jak i długie dystanse, a także miejsca,
nie dostępne np. dla transportu kolejowego lub morskiego. Wymienione cechy robią
drogowy przewóz towarów kluczowym elementem łańcucha dostaw, który zapewnia
efektywne funkcjonowanie całej gospodarki.

Jednocześnie, przewóz towarów wymaga precyzyjnego określenia ceny usługi. Jest to
związane z dużą konkurencją na rynku usług przewozów i ma istotny wpływ na relacje
pomiędzy uczestnikami rynku. Z drugiej strony, na ostateczną cenę usługi przewozowej
mają wpływ liczne czynniki ekonomiczne, geograficzne, polityczne, pogodowe etc. Wpływ
każdego z tych czynników na cenę końcową może być szacowany na podstawie danych
historycznych (np. popyt sezonowy), natomiast są też takie, których prognozowanie jest
bardzo trudne i wymaga dogłębnej analizy złożonych zależności.



Politechnika Warszawska

Wydział Transportu

Zagadnienie prognozowania cen na przewóz towarów jest więc bardzo ważnym elementem planowania działania operacyjnego firm transportowych oraz innych, powiązanych z nimi uczestników rynku. W ostatnich latach do tradycyjnych metod prognozowania cen na usługi przewozowe wykorzystujących podejścia statystyczne, empiryczne lub eksperckie, dołączyły metody bazujące na zastosowaniu technik uczenia maszynowego. Jest to związane między innymi z szybkim postępem technologicznym, co pozwoliło na efektywną realizację złożonych algorytmów z wykorzystaniem dedykowanego sprzętu (np. platforma CUDA firmy NVidia). Z kolei rozwój technologii cyfrowych (IoT, łączność, mapy cyfrowe etc.) umożliwił pozyskanie znacznie większej ilości danych, aniżeli jeszcze 10-15 lat temu.

Zastosowanie technik uczenia maszynowego nie jest nowością dla transportu. Jednocześnie należy wskazać, że obszar prognozowania cen na usługi transportowe wciąż pozostaje interesującym kierunkiem badań naukowych, przede wszystkim ze względu na różnorodność podejść do tematu oraz dużej liczby czynników wpływających na rezultat.

W tym kontekście należy więc stwierdzić, że tematyka badawcza rozprawy doktorskiej jest aktualna, dotyczy nowoczesnych technologii oraz ich zastosowania we współczesnym transporcie drogowym. Tematyka pracy wpisuje się w dyscyplinę naukową Inżynieria Lądowa, Geodezja i Transport.

2. Ogólna charakterystyka rozprawy

Praca składa się z 6 ponumerowanych rozdziałów napisanych w języku angielskim wraz z bibliografią (112 stron), streszczenia w języku angielskim (1 strona), streszczenia w języku polskim (1 strona), rozszerzonego streszczenia w języku angielskim (13 stron), rozszerzonego streszczenia w języku polskim (15 stron).

Rozdział pierwszy (5 stron) stanowi wprowadzenie do pracy. Rozdział 2 (3 strony) przedstawia autorską koncepcję badania. W Rozdziale 3 (17 stron) autor przedstawia podstawy teoretyczne do budowy modelu prognostycznych oraz uzasadnia wybór konkretnych modeli do zastosowania w badaniu. Rozdział 4 (36 stron) zawiera opis metod zbierania, transformacji oraz analizy danych, które autor wykorzystał w pracy. Kluczowy z punktu widzenia analizowanego zagadnienia badawczego rozdział 5 (43 strony) przedstawia wyniki obliczeń, wykonanych z wykorzystaniem zaproponowanej przez autora metody, a także ich interpretację. Rozdział 6 (3 strony) stanowi konkluzję pracy. Autor podsumowuje w nim wykonane zadania oraz osiągnięte wyniki, udowadniając tym samym poprawną weryfikację tezy rozprawy doktorskiej oraz osiągnięcie jej celu.

Sposób podania materiału w rozprawie doktorskiej należy uznać na poprawny. Objętość poszczególnych rozdziałów nie budzi zastrzeżeń. Praca jest napisana poprawnym stylistycznie językiem.



Politechnika Warszawska

Wydział Transportu

W pracy wykorzystano 88 pozycji literaturowych, wśród których znajdują się artykuły naukowe opublikowane w renomowanych czasopismach, takich jak Springer, Elsevier, MDPI oraz IEEE. Istotnym elementem bibliografii są pozycje poświęcone uczeniu maszynowemu, problemom transportu oraz organizacji przewozów towarów. Ponadto uwzględniono raporty branżowe i analizy rynkowe oraz akty prawne. Większość cytowanych źródeł pochodzi z ostatnich 5 - 7 lat, co świadczy o umiejętności skorzystania przez autora z aktualnego stanu wiedzy, szczególnie w dynamicznie rozwijającej się dziedzinie uczenia maszynowego.

Dobór literatury obejmuje wszystkie kluczowe aspekty pracy: problemy transportu drogowego towarowego, zagadnienia związane z ekonomicznymi aspektami kształtowania cen na przewóz, teorię uczenia maszynowego oraz przykłady jego zastosowania w prognozowaniu cen transportu. Istotnym elementem analizy literatury jest powiązanie problematyki rozprawy z rynkiem europejskim, co pozwala umiejscowić pracę w szerokim kontekście międzynarodowym. Uwzględnienie europejskich danych transportowych oraz realiów rynku FTL w UE dodatkowo wzmocnia merytoryczną wartość badań. W bibliografii znajdują się również pozycje odnoszące się do technologii implementacji modelu opracowanego przez autora - narzędzi i bibliotek języka Python. Dobór literatury w rozprawie jest ściśle związany z tematyką pracy, merytorycznie spójny oraz interdyscyplinarny. Odnosi się zarówno do aspektów teoretycznych, jak i praktycznego zastosowania metod predykcyjnych.

Rozdział 1 zawiera krótki przegląd badań naukowych w zakresie usług przewozu towarów, technik uczenia maszynowego jako narzędzia mającego na celu wspomaganie wydobywania wiedzy ze złożonych systemów transportowych, oraz uzasadnienie podjęcia się badań własnych w zakresie predykcji cen na usługi przewozowe. Na początku rozdziału autor podkreśla znaczenie transportu drogowego dla współczesnej gospodarki i logistyki, wykazuje aktualne trendy rozwoju usług przewozowych oraz podkreśla znaczenie technologii informatycznych w rozwiązaniu aktualnych problemów transportowych.

Następnie autor dokonuje przeglądu wybranych obszarów zastosowania technik uczenia maszynowego w transporcie, między innymi do zaawansowanej analizy dużych zbiorów danych, wspomaganie decyzji w systemach ruchu transportowego, optymalizacji procesów transportowych, zarządzania infrastrukturą drogową oraz prognozowania zapotrzebowania na usługi transportowe. Wymienia rodzaje modeli sztucznych sieci neuronowych stosowane w różnych dziedzinach transportu. Podkreśla przewagę, którą daje zastosowanie metod uczenia maszynowego w porównaniu z metodami klasycznymi. Krótko charakteryzuje narzędzia, które wspomagają proces tworzenia rozwiązań informatycznych wykorzystujących metody uczenia maszynowego, między innymi podstawowe środowisko badawcze współczesnego naukowca – Matlab, oraz język programowania Python jako wiodący obecnie język w zakresie praktycznych implementacji rozwiązań ML. Swoje wnioski autor popiera analizą bibliograficzną.



Politechnika Warszawska

Wydział Transportu

Dalej autor skupia się na zastosowaniu metod uczenia maszynowego w prognozowaniu cen na usługi transportowe. Podkreśla i uzasadnia istotność problemu dla branży transportowej, przede wszystkim z punktu widzenia ekonomicznego. Przytacza przykłady podejść klasycznych, opartych na statystycznej analizie danych. Wskazuje też na konieczność uwzględnienia w procesie prognozowania wielu aspektów, w tym pozatechnicznych, opartych na cechach przedsiębiorstw transportowych oraz relacjach między uczestnikami procesu. Następnie autor zwraca uwagę na efektywność integracji klasycznych metod ekonometrycznych z metodami opartymi na uczeniu maszynowym. Jest to uzasadnione i perspektywiczne podejście, ponieważ łączy funkcjonalność sprawdzonych modeli klasycznych z potęgą uczenia maszynowego w zakresie dogłębnej analizy powiązań oraz nieoczywistych zależności. Rozdział kończy podsumowanie ważności problemu predykcji cen usług przewozowych dla całej branży.

W opinii recenzenta, autor poprawnie i jednoznacznie uzasadnił obszar swoich badań w kontekście współczesnych problemów transportowych. Stwierdzam, że rozprawa doktorska prezentuje bardzo dobrą ogólną wiedzę teoretyczną autora w obszarach nauki związanych z transportem towarów oraz powiązanymi zagadnieniami z zakresu metod uczenia maszynowego.

3. Ocena merytoryczna rozprawy

Na początku rozdziału drugiego autor na podstawie wykonanej analizy literatury uzasadnia konieczność przeprowadzenia własnego badania ze względu na coraz większą złożoność procesu prognozowania cen na usługi transportowe i wynikającą z tego potrzebę zastosowania coraz bardziej zaawansowanych metod i narzędzi informatycznych. Następnie definiuje lukę badawczą polegającą na braku kompleksowego podejścia do budowy modeli uczenia maszynowego w segmencie prognozowania cen na usługi transportowe w transporcie drogowym. Podkreśla konieczność opracowania jednolitych ram, które pozwolą na sterowanie całym cyklem życia modelu, co prowadzić będzie do przewidywalnych i powtarzalnych wyników.

Następnie autor formułuje problem badawczy polegający na trudności prognozowania cen drogowego transportu towarowego ze względu na dużą liczbę współzależnych zmiennych i czasochłonny charakter procesu prognozowania. Stawia pytanie badawcze polegające na identyfikacji kluczowych czynników mających wpływ na kształtowanie cen na usługi transportowe, opracowaniu metodyki zbierania i przetwarzania danych wejściowych pochodzących z przedsiębiorstw transportowych do nauczania sieci neuronowej, oraz porównaniu efektywności podejścia autora z metodami ocen eksperckich.

Główny cel pracy autor definiuje jako opracowanie solidnej metodologii prognozowania cen usług transportu drogowego towarów z wykorzystaniem technik uczenia maszynowego.



Politechnika Warszawska

Wydział Transportu

Dalej autor wymienia zadania, które należy zrealizować na drodze do osiągnięcia celu. W zakończeniu rozdziału autor formułuje tezę rozprawy, w której stawia hipotezę roboczą, że wdrożenie technik uczenia maszynowego do prognozowania cen usług transportu drogowego poprawi dokładność prognoz oraz umożliwi firmom transportowym dostosowanie decyzji operacyjnych w odpowiedzi na zmieniające się warunki. Wymienia też oczekiwane korzyści ze wdrożenia takiego rozwiązania w praktyce.

W rozdziale 3 autor wprowadza najważniejsze podstawy teoretyczne do budowy modelu prognostycznych. Opisuje ich charakterystyki i uzasadnia wybór konkretnych modeli. Przedstawia cykl życia modelu. Wymienia narzędzia informatyczne wykorzystane w pracy. Uzasadnia wybór języka Python do programowej realizacji modułu uczenia maszynowego. Charakteryzuje biblioteki języka Python wykorzystane w celu budowy modeli, manipulacji danymi oraz wizualizacji rezultatów badań. Deklaruje średni błąd procentowy (MPE) jako kryterium porównania dokładności predykcji. Podkreśla znaczenie walidacji krzyżowej jako jednej z podstawowych technik stosowanych w uczeniu maszynowym do oceny oraz optymalizacji modelu. Rozpatruje wybrane metody prognozowania w uczeniu maszynowym, takie jak metoda średniej historycznej, regresja liniowa oraz jej modyfikacje w postaci regresji grzbietowej i regresji z regularyzacją, drzewa decyzyjne oraz ich modyfikacje. Przedstawia zalety i wady każdej z metod oraz przykłady dziedzin ich zastosowania. Opisuje algorytmy regresji AdaBoost oraz XGBoost oraz ich przydatność do rozwiązania zadań prognozowania cen usług transportowych. Kończy rozdział krótkie omówienie zastosowania oceny eksperckiej w celu uzupełnienia oraz walidacji modeli uczenia maszynowego.

W rozdziale 4 autor skupia się na metodach zbierania, transformacji oraz analizy danych, które wykorzystuje do budowy modelu prognostycznego. Na początku rozdziału autor przedstawia metodę rejestrowania danych wejściowych, podkreślając ważność ich jakości do uzyskania jak najlepszego rezultatu prognoz. Dokonuje wyboru formatu danych wejściowych CSV i uzasadnia swój wybór. Następnie omawia wpływ jakości kodu źródłowego na rozwój i wsparcie produktu programowego. Proponuje wprowadzenie regionalizacji przewozów i uwzględnienia specyfiki kształtowania cen w zależności od kraju. Wprowadza klasyfikację takich cech usługi przewozowej jako: odległość, relacja, data, rodzaj ładunku, rodzaj transportu oraz inne cechy. Definiuje format zapisu każdej z wymienionych cech.

Następnie autor omawia kwestie przygotowania danych wejściowych do modelu predykcyjnego. Przedstawia metody transformacji danych: standaryzację, czyszczenie, agregację oraz integrację. Standaryzacja danych polega na ich dostosowaniu do formatów zapisów, przedstawionych wcześniej. Czyszczenie danych polega na odrzuceniu danych nieprawidłowych oraz ich zastąpieniu wartością domyślną, wygenerowaną np. wg wzoru. Agregacja danych jest wykorzystywana do podsumowania trendów zmian w wybranych okresach. Integracja danych polega na wykorzystaniu danych pochodzących z różnych źródeł, oraz ich doprowadzenie do wspólnego ujednoliconego zestawu. Dalej autor



Politechnika Warszawska

Wydział Transportu

dokonyuje szczegółowego opisu transformacji danych w systemie. Należy podkreślić, że oprócz gromadzenia danych autor przewiduje w modelu ich wstępną statystyczną obróbkę oraz umieszczenie jej wyników w polach struktur danych. Takie podejście pozwoli na lepszą interpretację obserwowanych zależności. W procesie przetwarzania danych autor wskazuje na dodatkowe cechy opisujące proces ładunku i rozładunku: średnią, medianę i odchylenie standardowe. Ich umieszczenie w modelu pozwoli na dokładniejsze spojrzenie w głąb analizowanych procesów.

W zakończeniu rozdziału 4, autor określa kategorie danych, które będą stanowiły podstawę do analizy zależności ceny usługi przewozowej od poszczególnych czynników. Jest to kluczowy fragment do zrozumienia procesu kształtowania ceny końcowej oraz niezbędny czynnik dokładności jej prognozowania. Wyjaśnia znaczenie korelacji pomiędzy ceną na przewóz a połączeniami geograficznymi, porami roku, cenami na paliwo etc. Korzysta z zewnętrznych baz danych do uzupełnienia modelu o dane, które byłyby trudne w pozyskaniu we własnym zakresie.

W opinii recenzenta, autor rozprawy doktorskiej poprawnie przeprowadził badania wstępne, ujawnił lukę badawczą oraz sformułował cel pracy. Przedstawił rzetelny oraz szczegółowo opisany plan wykonania poszczególnych kroków, niezbędnych do osiągnięcia celu:

- przegląd i dobór modeli prognostycznych, które będą wykorzystane w badaniu,
- określenie kryteriów jakości prognozowania,
- analiza czynników wpływających na cenę usługi przewozowej,
- definiowanie zmiennych modelu oraz analiza ich współzależności,
- gromadzenie i uporządkowanie danych źródłowych do analizy,
- przygotowanie zbiorów walidacyjnych oraz opinii eksperckich na potrzeby sprawdzenia poprawności oraz dokładności modelu,
- klasyfikacja danych oraz opracowanie formatów danych,
- przeprowadzenie eksperymentu, analiza oraz interpretacja jego wyników,
- weryfikacja hipotezy roboczej oraz udowodnienie osiągnięcia celu pracy.

Na tej podstawie stwierdzam, że w rozprawa doktorska wykazuje umiejętność samodzielnego prowadzenia pracy naukowej przez autora na wysokim poziomie.

W rozdziale 5 autor przedstawia rezultaty predykcji uzyskane z wykorzystaniem własnej metody. Na początku podaje charakterystyki statystyczne opisujące zbiór danych wejściowych. Następnie wykorzystując współczynnik korelacji Pearsona, analizuje korelację pomiędzy wybranymi cechami numerycznymi modelu. Dalej szczegółowo omawia wykryte zależności, ilustrując to w formie graficznej za pomocą wykresów. Wyjaśnia charakter poszczególnych zależności i daje im własną interpretację. Wskazuje na asymetryczność kosztów w usługach transgranicznych. Pokazuje wzrost popytu na usługi przewozowe dla poszczególnych miesięcy w roku oraz powtarzalność tych trendów na przestrzeni kilku lat. Zwraca uwagę na wpływ wydarzeń nadzwyczajnych na kształtowanie ceny przewozu. Analizuje udział pojazdów różnego typu w przewozach, w zależności od odległości pomiędzy



Politechnika Warszawska

Wydział Transportu

punktami ładunku oraz rozładunku. Omawia wpływ kosztów dodatkowych oraz usług celnych na końcowy koszt przewozu.

Następnie autor porównuje dokładność prognozowania ceny usług przewozowych dla różnych modeli regresji, na tym samym zbiorze danych. Odrzuca modele, które przekroczyły akceptowalny poziom średniego bezwzględnego błędu procentowego MAPE (10%). Uzasadnia zastosowanie wzajemnej walidacji wyników prognozowania ceny usług przewozowych poprzez metody walidacji krzyżowej oraz podziału opartego na czasie. Dalej skupia się na zagadnieniu wpływu rozmiaru zbioru danych treningowych na dokładność prognozowania cen. Przedstawia graficznie zależności MAPE od liczby rekordów w zbiorze treningowym. Pokazuje punkty nasycenia dla poszczególnych modeli ML.

Ważnym elementem pracy jest dokonane przez autora zestawienie wyników uzyskanych drogą modelowania, z ocenami eksperckimi. W dalszej części rozdziału autor szczegółowo przedstawia grono ekspertów i charakteryzuje go pod względem wieku, doświadczenia, rozmiaru i rodzaju działalności przedsiębiorstwa, etc. Przedstawia rezultat badań w postaci zbioru danych walidacyjnego, który składa się z 5 ofert transportowych. Wykorzystuje rezultaty ocen eksperckich jako dane walidacyjne do prognoz uzyskanych z modeli predykcyjnych. Wykazuje przewagę modeli maszynowych nad oceną ekspercką.

Następnie autor analizuje korelacje pomiędzy ceną paliwa jako najbardziej znaczącym składnikiem ceny usługi przewozowej, a stawką za 1 km przewiezienia ładunku. Pokazuje, że dodanie do modelu wskaźnika PKB kraju nie zwiększa dokładności prognozowania. Analizuje też pozostałe składowe ceny końcowej. Przedstawia tabelę podsumowującą wpływ poszczególnych czynników na wielkość błędu MAPE. Porównuje czas nauczania modelu dla różnych modeli oraz konfiguracji czynników. Następnie, autor przechodzi do omówienia najważniejszych cech modelu predykcji oraz nadanych im wag. W ten sposób autor finalizuje proces przedstawienia charakterystyk własnego modelu predykcyjnego.

Podsumowując ocenę merytoryczną pracy, należy stwierdzić, że:

- autor poprawnie przeprowadził badanie, opracował oryginalną metodę prognozowania cen na usługi transportowe oraz pozyskał wyniki, prawidłowość których jest potwierdzona przez walidację modelu, zbieżna z ocenami eksperckimi oraz przewyższa je dokładnością prognoz uzyskanych na identycznym zbiorze danych wejściowych; potwierdza to słuszność hipotezy postawionej przez autora na początku pracy,
- autor poprawnie określił czynniki, które wpływają na wycenę usługi transportowej, wskazał również ich wpływ na rezultat prognoz; należy tu podkreślić, że jedynie w przypadku długości trasy przejazdu ta zależność jest dość oczywista – wpływ pozostałych cech na cenę końcową stanowi oryginalny rezultat pracy autora,
- zbudowany przez autora model wykazuje się szybkością działania, która pozwala wykorzystać go w zastosowaniach komercyjnych np. jako część komputerowego systemu logistycznego; realizacja modelu w Python znacząco ułatwia w perspektywie integrację kodu źródłowego,



Politechnika Warszawska

Wydział Transportu

- autor udowodnił eksperymentalnie, że włączenie do modelu predykcji danych z zewnętrznych źródeł (bazy danych cen paliw, opłaty drogowe etc.) dodatkowo zwiększa dokładność prognoz,
- model predykcji opracowany przez autora może być przeniesiony na inne rynki niż europejskie; może być również łatwo rozszerzony o dodatkowe czynniki wpływające na cenę końcową usługi przewozowej, co może doprowadzić do jeszcze większej dokładności prognozowania; model może być również adaptowany dla alternatywnych źródeł paliwa.

W związku z powyższym stwierdzam, że przedstawiona do oceny rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego, jakim jest prognozowanie cen usług przewozowych w transporcie drogowym.

4. Uwagi i dyskusja

4.1. Uwagi merytoryczne

1) Autor na przemian stosuje pojęcia „metodologia” oraz „metoda”. Przykładowo, w 2.3 planuje opracować całą metodologię, natomiast w rezultatach pracy w rozdziale 6 wskazuje, że opracowana została jednak metoda. Metodologię też można spotkać w nagłówkach rozdziałów. W opinii recenzenta, nie możemy mówić o metodologii w przypadku osiągnięć autora. Oznaczałoby to kompletne opracowanie nowego podejścia do rozwiązywania problemu transportowego. Natomiast w treści została przedstawiona oryginalna metoda autorska.

2) W podrozdziale 3.2 w pierwszym akapicie autor pisze, że rozdział skupia się na „wyborze odpowiednich technik uczenia maszynowego ...”, natomiast w części dalszej pisze o Pythonie, bibliotekach, środowisku Jupyter etc. Dopiero na końcu podrozdziału można znaleźć wzmiankę o algorytmie XGBoost jako „doskonałym wyborze do prognozowania cen ...”, lecz w żaden sposób wniosek ten nie został poparty faktami – ani porównaniem z innymi algorytmami, ani cytataми z bibliografii. Dlaczego tylko ten algorytm zasłużył na szczegółowy opis? Dowiemy się dopiero w podrozdziale 3.5.

3) Praca łączy zagadnienia transportowe oraz informatyczne. Część informatyczna jest bardzo istotna do oceny osiągnięcia autora. W środowisku informatycznym przyjęto jest stosowanie graficznych ilustracji w celu opisanego struktury systemu, zależności pomiędzy obiektami, procesów zachodzących w systemie etc. (np. w postaci diagramów UML). W pracy brakuje niestety takiego typu wizualizacji. Przydałaby się ona przede wszystkim w opisie źródeł i formatów danych wejściowych, ich relacji, algorytmów transformacji, a na etapie modelowania – w postaci schematu co jest na wejściu oraz na wyjściu systemu. Znacząco ułatwiłoby to czytanie pracy.

4) W pracy naukowej nie należy używać epitetów w stylu „jest idealnym językiem dla początkujących”, „jest doskonałym narzędziem”, „wyróżnia się jako przydatny”, „cenny wybór dla badaczy”, „drastycznie skrócił czas” etc. Sformułowania te są zbyt emocjonalne jak na pracę naukową, poza tym są subiektywne. Natomiast odnośnie uznania Pythona za idealny język dla początkujących też mam wątpliwość. Python jest współczesnym, modnym,



Politechnika Warszawska

Wydział Transportu

dobrze wyposażonym w biblioteki językiem programowania. Można powiedzieć, że dla doświadczonych programistów jest to dobry kierunek rozwoju. Natomiast do nauki programowania od podstaw nie jest to wybór idealny chociażby ze względu na specyficzną składnię oraz dynamiczne typy danych.

5) Niektóre fragmenty w pracy można byłoby spokojnie pominąć jako nie wpływające na jej odbiór. Przykładowo, wzmiankę o niezgodności pomiędzy Python 2 a Python 3, lub zbyt dokładny opis środowiska Jupyter.

6) Przykłady efektywności wybranych metod uczenia maszynowego generalnie pochodzą spoza tematyki transportowej; a tam, gdzie mieszczą się w dziedzinie, można spotkać np. „model osiągnął wyniki wariancji ... odpowiednio dla kątów natarcia 4 st. i 8 st.”; trudno się odnieść do takiego wyniku, ponieważ dla nie specjalisty z dziedziny aerodynamiki te kąty nic nie mówią.

7) Pisząc o cross-validation, autor wspomina między innymi o tym, że “that is used to evaluate model performance, optimize hyperparameters”. Niestety, w dalszej części pracy nie znalazłem opisu hiperparametrów zastosowanych w modelu autorskim.

8) W podrozdziale 3.5. (metody prognozowania) autor przedstawia słowny opis algorytmów i metod. O przydatności ilustracji graficznych już pisałem wyżej. Dodam w tym miejscu, że przydałaby się w tym miejscu tabela z podsumowaniem porównywanych metod i algorytmów, lub wykres ilustrujący ich osiągnięcia.

9) W rozdziale 4 autor nawiązuje do technik pozyskiwania i przetwarzania danych, znanych w informatyce pod nazwą ETL (extract-transform-loading) i szeroko stosowanych na przykład w zasilaniu hurtowni danych. Sam ten skrót jednak w pracy nie występuje. Poza tym, nazwa rozdziału powinna bardziej nawiązywać do przetwarzania danych.

10) Można dyskutować, czy format CSV jest właściwym formatem danych wejściowych, przede wszystkim ze względu na istnienie wielu „dialektów” oraz brak jednolitej standaryzacji. Wykorzystanie tego formatu do przechowywania danych pochodzących z różnych krajów wymaga zastosowania ścisłej weryfikacji jakości danych wejściowych (przykładowo, formatu liczb rzeczywistych). Z pewnością CSV jest formatem prostym w obsłudze, dlatego nie neguję wyboru autora. Chciałbym jedynie podkreślić, że programowa realizacja algorytmów modelu zwłaszcza w środowiskach wykonania o rozmaitej regionalizacji (język, daty, liczby, etc.) wymaga obsługi różnorodnych formatów.

11) W rozdziale 4.1 jest wstawka dotycząca zalet „czystego kodu”. Zawiera ona między innymi stwierdzenie, że „clean code plays a key role in effectively conducting analyses and developing predictive models”. Nie do końca mogę się z tym zgodzić, że jest to kluczowy element sukcesu. Raczej oczywistym jest że utrzymanie kodu zgodnie z przyjętymi zasadami sprzyja jego rozwojowi oraz utrzymaniu. Ale te aspekty wykraczają poza ramy rozprawy. Wymieniając zalety „czystego kodu”, autor niestety nie wskazuje, na czym to podejście miałoby polegać w jego pracy – poza „proponuje się używanie języka angielskiego do nazewnictwa funkcji”. Co moim zdaniem w pracy napisanej w języku angielskim, jest oczywiste. Dlatego uważam, że fragment ten mógłby być znacznie krótszy.

12) W podrozdziale 4.3 o nazwie „Data Analysis Methods” autor przedstawia nie tyle metody analizy danych, ile czynniki mające wpływ na kształtowanie ceny przewozu towarów



Politechnika Warszawska

Wydział Transportu

(cena paliwa, odległość, pora roku, rodzaj transportu itp.) oraz zachodzące pomiędzy nimi zależności wyjaśnione za pomocą korelacji.

13) W rozdziale 5 autor nie określił dokładnego pochodzenia zbiorów danych, na podstawie których przeprowadził analizę. Ogranicza się do stwierdzenia „using a real-word dataset”. Poza tym, powołanie się na rezultat funkcji `df.info()` jest zbędne i wygląda nieco dziwnie. Struktura DataFrame nie zależy od tej funkcji.

14) Wykres 6 na stronie 65 zaczyna się od dystansu 0 km, dla którego cena jednostkowa przewozu sięga powyżej 7 euro. Z czego wynika ten koszt? Jak należy to interpretować?

15) W tabeli 28 na stronie 66 sumaryczny kilometrów tras na terenie Polski (PL_KM) w porównaniu z pozostałymi krajami jest bardzo wysoki. Z czego to wynika? Czy analizowany był zbiór danych dla firm logistycznych, czyja siedziba (lub rejestracja pojazdów) znajduje się w Polsce? Analogiczne pytanie dotyczy liczby transportów.

16) Rozdział 5.6 ma nazwę „Implementation of External Databases”, która może sugerować, że bazy zewnętrzne będą przez autora w jakiś sposób realizowane. W podrozdziale natomiast chodzi o wykorzystanie danych z zewnętrznych baz danych w celu zwiększenia dokładności prognoz.

17) Tab. 45 przedstawia czas treningu dla różnych modeli. Jak można zauważyć, ten sam zbiór danych dla jednego z algorytmów może wydłużyć czas nauczania, a dla drugiego – skrócić (np. GBoost i XGBoost, kolumny basic vs fuel). Czy w przypadku innego wejściowego zbioru danych treningowego, można spodziewać się że proporcje w czasie nauczania dla poszczególnych modeli zostaną zachowane?

4.2. Uwagi redakcyjne

Str. 1 (spis treści) literówka w 5.4 „performance”

Str. 29 – „despite the Polish language in the doctoral dissertation ?!

Str. 31 – “Tab 2 presents the proposed method .. “ – it is format, not method

Str. 33 – Tab. 4 “Date gathering method” – powinno być raczej Date gathering format ?

Str. 40 – Tab. 9 literówka w nagłówku pierwszej kolumny „VIHECLE”, powinno być VEHICLE

Str. 40 – kod VEHICLE_TYPE zawiera identyfikatory typu „<12.5t”. W rozdziale 3 mówiliśmy o jakości kodu – to nie jest dobry przykład nazewnictwa enumeratora dla typu pojazdu

Str. 48 – tabelę należy podnieść wyżej, wówczas tekst „equation 7 presents” będzie poprzedzał formułę a nie tabelę, co dziwnie wygląda,

Str. 48 – niewykorzystane miejsce na stronie

Str. 61 – Tab. 26 można by było przedstawić w bardziej kompaktowej postaci (np. w dwóch sekcjach oraz mniejszą czcionką)

Str. 61, 62, 66 – trzy tabele mają różny rozmiar czcionki

Str. 62 – niewłaściwe formatowanie akapitów

Str. 62 - 67 (dalej również, co kilka stron) za dużo niewykorzystanego miejsca na stronach, można by było przenieść fragmenty tekstu w taki sposób, aby wykorzystać powierzchnię strony maksymalnie

Str. 74 nadużycie czcionki pogrubionej w pierwszej kolumnie tabeli

Str. 86 urwany zapis Tab.16 ??



Politechnika Warszawska

Wydział Transportu

Rozszerzone streszczenie w języku polskim

Str. 1 – tytuł pracy w języku angielskim

Str. 9 – nazwa rozdziału w języku angielskim

Str. 13 Extended abstract – część zdań jest w języku polskim

5. Wniosek końcowy

Temat rozprawy doktorskiej jest ściśle związany z Dyscypliną Naukową Inżynieria Lądowa, Geodezja i Transport. Praca dotyczy aktualnych zagadnień związanych z predykcją cen na usługi przewozów towarowych z wykorzystaniem technik uczenia maszynowego. Autor wykazuje się zdolnością do prowadzenia samodzielnej pracy badawczej. Układ pracy jest prawidłowy. Uwagi wymienione w pkt. 4 nie wpływają na ogólną pozytywną ocenę pracy przez recenzenta.

W związku z powyższym stwierdzam, że rozprawa doktorska mgr inż. Artura Budzyńskiego pt. „Forecasting prices for road freight transport services using machine learning” spełnia wymogi Ustawy z dnia 20 lipca 2018 r. (z późniejszymi zmianami) „Prawo o szkolnictwie wyższym”. Wniosuję do Rady Naukowej Dyscypliny Inżynieria Lądowa, Geodezja i Transport Politechniki Śląskiej o dopuszczenie przedmiotowej rozprawy do obrony publicznej.

Z wyrazami szacunku,

Andrzej Czerep/cki

