

POLITECHNIKA ŚLĄSKA
WYDZIAŁ AUTOMATYKI, ELEKTRONIKI I INFORMATYKI

ROZSZERZONE STRESZCZENIE ROZPRAWY DOKTORSKIEJ

**Solutions for selected problems of
demand forecasting based on machine
learning methods and domain
knowledge**

**(Rozwiązania wybranych problemów
prognozowania zapotrzebowania w
oparciu o metody uczenia
maszynowego i wiedzę dziedzinową)**

Joanna Badura

Promotor: dr hab. Marek Sikora, prof. Politechniki Śląskiej

Gliwice 2023

1 Wstęp

Rozprawa doktorska skupia się na wykorzystaniu metod uczenia maszynowego i wiedzy dziedzinowej do rozwiązywania wybranych zagadnień związanych z prognozowaniem zapotrzebowania.

Zrozumienie i dokładne prognozowanie zapotrzebowania na produkty, usługi i media ma kluczowe znaczenie dla firm. Wpływa to na zarządzanie zasobami, wydajność łańcucha dostaw i na zadowolenie klientów. Metody oparte na prognozowaniu osądowym (*judgmental forecasting* [18]) i na decyzjach ekspertów są nadal stosowane, ale coraz częściej, zwłaszcza przy bardziej skomplikowanych problemach, do tego zadania wprowadzane są metody sztucznej inteligencji. Uczenie maszynowe, poddziedzina sztucznej inteligencji, oferuje wiele metod, które można wykorzystać do prognozowania zapotrzebowania. Są one przydatne w przypadkach, gdy na prognozowaną wartość wpływa wiele czynników, a wzorce wpływające na popyt są bardzo złożone.

Myśląc o prognozowaniu zapotrzebowania, pierwszą rzeczą, która może przyjść do głowy, jest problem prognozowania sprzedaży produktu, który był już w sprzedaży, czyli dla którego mamy dane historyczne. W tym przypadku sprzedaż danego produktu może być przedstawiona w formie szeregu czasowego. Kwestie prognozowania popytu nie ograniczają się jednak wyłącznie do danego zagadnienia. Temat ten jest znacznie szerszy, a dana praca doktorska poświęca uwagę różnym aspektom związanym z prognozowaniem zapotrzebowania. Niektóre z poruszanych aspektów zostały zidentyfikowane jako luka badawcza w wykorzystaniu metod uczenia maszynowego.

W rozprawie przedstawiono autorską taksonomię zagadnień związanych z prognozowaniem zapotrzebowania. Następnie wybrane zagadnienia zostały omówione bardziej szczegółowo. W badaniach wykorzystano rzeczywiste dane z przedsiębiorstw należących do różnych branż. Często są to zbiory obejmujące kilka lat i dotyczące wielu produktów. Ze względu na ich wrażliwy charakter, większość z nich jest zastrzeżona.

Autorka rozprawy doktorskiej jest stypendystą Unii Europejskiej w ramach Europejskiego Funduszu Społecznego, grant InterPOWER (POWR.03.05.00-00-Z305).

1.1 Teza pracy

Dzięki rozbiciu zadania prognozowania na konkretne zagadnienia i dzięki uwzględnieniu wiedzy dziedzinowej można uzyskać użyteczne wartości prognoz za pomocą metod uczenia maszynowego.

Teza pomocnicza

Wiedza dziedzinowa może zostać uwzględniona poprzez wprowadzenie do modelu nowych atrybutów lub nowych rekordów danych z podobnych problemów, co skutkuje poprawą ostatecznych wyników prognozy.

1.2 Cele pracy

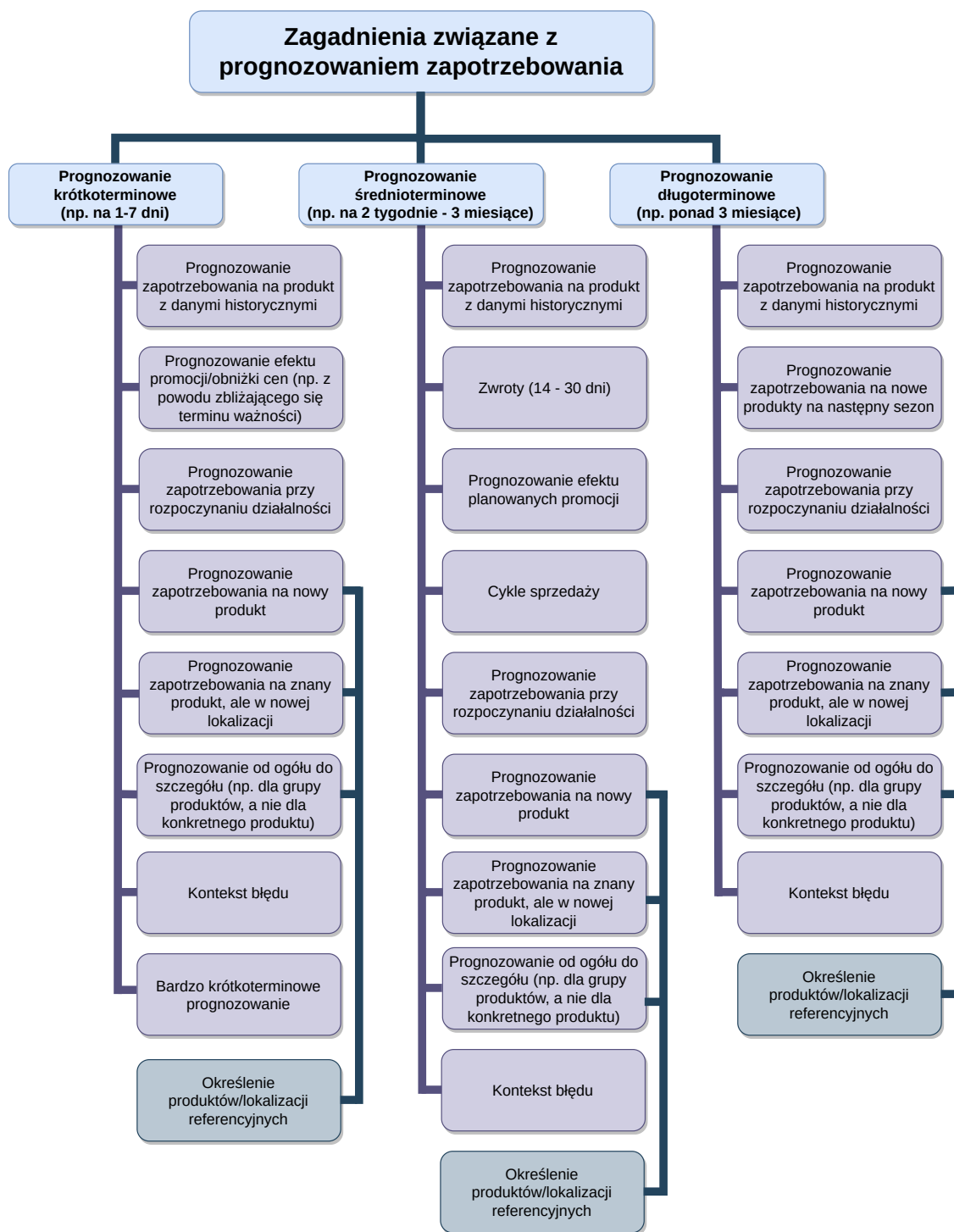
Pierwszym celem rozprawy było zaproponowanie taksonomii zagadnień związanych z prognozowaniem popytu, które mogą być rozwiązane metodami uczenia maszynowego. Kolejnym celem było zaproponowanie wykorzystania metod uczenia maszynowego i wiedzy dziedzinowej do rozwiązania wybranych zagadnień. Wybór tych zagadnień został zdeterminowany przez przeprowadzone projekty badawcze.

1.3 Układ pracy

Praca doktorska składa się z 7 rozdziałów. Rozdziały 1 i 7 stanowią odpowiednio wprowadzenie i podsumowanie. Rozdział 2 przedstawia wprowadzenie do tematu prognozowania zapotrzebowania z wykorzystaniem uczenia maszynowego i wiedzy dziedzinowej. W rozdziale 3 przedstawiono temat prognozowania efektywności promocji. Rozdział 4 koncentruje się na problemie prognozowania od ogółu do szczegółu (*top-down forecasting*). Rozdział 5 przedstawia zagadnienie wykorzystania dodatkowych szeregów czasowych w celu poprawy prognozowania badanego zjawiska. Rozdział 6 koncentruje się na prognozowaniu prawdopodobieństwa zwrotu produktu.

2 Prognozowanie zapotrzebowania – podejścia i wyzwania

W rozprawie przedstawiono autorską taksonomię zagadnień związanych z prognozowaniem zapotrzebowania. Zostały one podzielone na trzy kategorie: przypadki związane z prognozowaniem krótkoterminowym, średnioterminowym oraz długoterminowym. Ta taksonomia jest podyktowana różnymi przypadkami biznesowymi. Każde zagadnienie zostało opisane w pracy dyplomowej i omówiono problemy, które mogą się z nim wiązać. Rysunek 1 przedstawia zaproponowany w niniejszej pracy nowatorski podział zadań związanych z prognozowaniem zapotrzebowania. Następnie wybrane zagadnienia zostały wybrane do dalszych badań.



Rysunek 1: Taksonomia zagadnień związanych z prognozowaniem zapotrzebowania

3 Promocje

Pierwszym tematem wybranym do dalszej analizy był temat prognozowania skuteczności promocji. W pierwszym etapie badań zaproponowano sześć wskaźników, których celem było opisanie efektywności promocji [11]. Następnie przeprowadzono badania nad modelowaniem danych wskaźników. Proces przygotowania danych treningowych i inżynierii cech został przedstawiony w pracy doktorskiej. W pierwszym podejściu modele prognostyczne dla wskaźników zostały utworzone na podstawie danych historycznych wszystkich produktów z danej grupy produktów, a nie dla poszczególnych produktów. Przeprowadzono porównanie modelu XGBoost (eXtreme Gradient Boosting) [5], sekwencyjnych modeli głębokiego uczenia i modeli głębokiego uczenia z pakietu H2O. Biorąc pod uwagę miarę, która była optymalizowana, najlepsze wyniki uzyskano dla XGBoost w 10 z 12 przypadków. Badania zostały również zaprezentowane w artykule [10].

W kolejnym podejściu przetestowano możliwość prognozowania wybranych wskaźników na podstawie danych historycznych dotyczących wyłącznie badanego produktu [2]. Przeprowadzono porównanie algorytmów z 4 rodzin algorytmów. Analiza wartości korelacji (R) została przeprowadzona dla każdej metody i każdego badanego produktu. Wartość R była zazwyczaj wysoka, jednak w przypadku niektórych produktów wskaźnik ten spadł nawet do 0,64. Zbadano niepokojącą dużą wariancję wyników i okazało się, że korelacja R modeli koreluje z wielkością zbioru danych historycznych. Potwierdziło to pierwotne założenie, że dla wybranych produktów, dla których istnieje wiele danych o historycznych promocjach, prognozowanie efektu promocji w oparciu tylko o dane historyczne danego produktu może być możliwe. Jednak w przypadku wielu produktów takich danych może być zbyt mało, a w takim przypadku przydatne może być prognozowanie efektu promocji z uwzględnieniem danych historycznych dotyczących wielu produktów z tej samej grupy produktowej.

Następnie zaprezentowano praktyczne zastosowanie stworzonych modeli prognostycznych. W pierwszej iteracji celem było wykorzystanie wcześniej przygotowanych danych i modeli prognostycznych w celu stworzenia prototypu symulatora wrażliwości cenowej – rozwiązania do modelowania wpływu promocji (zmiany ceny) na sprzedaż i zysk. W zaproponowanym rozwiązaniu użytkownik może wskazać parametry przyszłej promocji, a na wyjściu otrzymuje informacje o tym, jak zmienia się sprzedaż produktu i zysk dla różnych wartości cenowych. Celem drugiej iteracji było zaproponowanie prototypu systemu rekomendacji promocyjnych. System wskazywał cechy

najkorzystniejszej promocji, czyli takiej, która dla danych parametrów wejściowych uzyskiwała najwyższą prognozę dla wskaźnika opisującego średnią liczbę sprzedanych sztuk lub kilogramów każdego dnia danego produktu. Wśród parametrów wejściowych znalazły się ograniczenia, które kontrolowały wartości innych wskaźników dla rekomendowanej promocji. Takie rozwiązanie umożliwiło znalezienie promocji, która jest odpowiednia, biorąc pod uwagę kilka wskaźników jednocześnie. W rozprawie przedstawiono interfejs oraz logikę prezentowanych rozwiązań.

W ostatniej części rozdziału poświęconego promocji przedstawiono możliwość zastosowania nowatorskiej metody indukcji przeżyciowych reguł akcji (Survival Action Rules) [1] w celu znalezienia zmian, które mogą poprawić efekt promocji. W rozprawie przedstawiono teorię dotyczącą analizy przeżycia [21, 24], indukcji reguł [20] oraz reguł akcji [6]. Przeżyciowe reguły akcji są połączeniem typowych reguł akcji generowanych dla problemów klasyfikacji i reguł przeżycia. Ten rodzaj reguły wskazuje działanie, które należy wykonać (opisane w przesłance) w celu zmiany krzywej przeżycia (która jest zawarta w konkluzji reguły). Zaproponowany algorytm indukcji przeżyciowych reguł akcji jest pokryciowym algorytmem i składa się z dwóch etapów: wzrostu i przycinania reguły. Szczegóły omawianego algorytmu, wraz z pseudokodami, są zawarte w rozprawie doktorskiej.

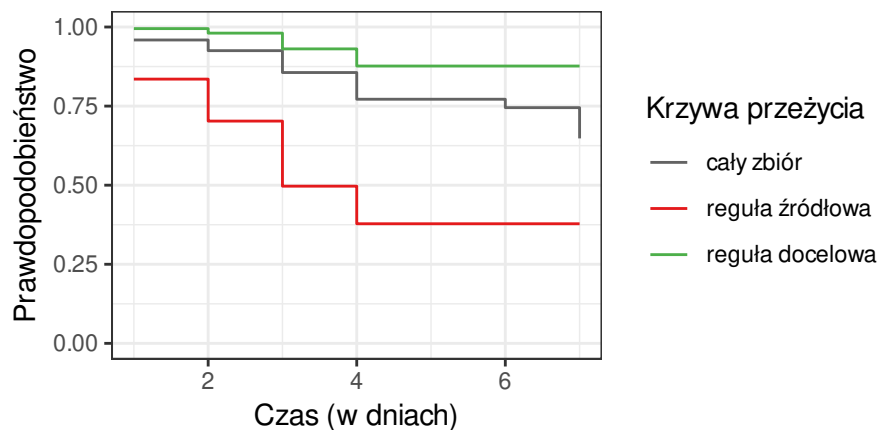
W przeprowadzonym badaniu rozpatrywanym zdarzeniem analizy przeżycia był spadek wartości wskaźnika opisującego promocję poniżej ustalonego progu. Czas przeżycia opisywał dzień promocji, w którym zdarzenie miało miejsce lub, w przypadku braku takiego zdarzenia, czas trwania promocji. Taka reprezentacja pozwoliła na zastosowanie metody indukcji przeżyciowych reguł akcji. Wynikowy zestaw reguł pokazał zmiany (akcje) w parametrach promocji, które zmniejszają prawdopodobieństwo spadku wskaźnika poniżej danego progu w czasie trwania promocji. Wszystkie z uzyskanych dziecięciu reguł miały p-wartość testu log-rank mniejszą niż 0,01, co wskazywało, że były one dobrej jakości.

Studium przypadku wykorzystania przeżyciowych reguł akcji dla tematu promocji zostało przedstawione na przykładzie jednej wybranej reguły. Przesłankę danej reguły pokazano w Tabeli 1, a krzywe przeżycia dla reguły źródłowej i docelowej przedstawiono na Rysunku 2. Przesłanka reguły składa się z pięciu warunków, w tym czterech akcji i jednego warunku elementarnego. Akcje zostały otrzymane dla atrybutów: liczba wszystkich promocji w sklepie, które były reklamowane w telewizji, radiu lub Internecie (nb_promos_TRI), liczba wszystkich promocji, które były reklamowane w inny sposób (nb_promos_D), cena ($price$) i liczba wszystkich promocji

w sklepie (nb_promos). Przesłanka reguły pokazuje, jakie zmiany powinny nastąpić w wartościach atrybutów, aby zmienić krzywą przeżycia dla badanego wskaźnika (przedstawione na Rysunku 2). Algorytm indukcji przeżyciowych reguł akcji i przedstawione studium przypadku pokazały nowy sposób rekomendowania zmian, które można zastosować, aby promocja spełniała oczekiwaną wartość wskaźnika przez dłuższy czas.

Premise
$(nb_promos_TRI < 18 \rightarrow nb_promos_TRI > 2) \text{ AND}$
$(nb_promos_D < 118 \rightarrow nb_promos_D > 22) \text{ AND}$
$(price < 1.57 \rightarrow price > 1.64) \text{ AND } (first_date_day_of_year < 339) \text{ AND}$
$(nb_promos \in <37, 302) \rightarrow nb_promos > 143)$

Tabela 1: Przesłanka analizowanej przeżyciowej reguły akcji.



Rysunek 2: Krzywe przeżycia konkluzji reguły, której przesłanka została przedstawiona w tabeli 1. Krzywa odpowiadająca regule źródłowej jest oznaczona kolorem czerwonym, krzywa odpowiadająca regule docelowej jest oznaczona kolorem zielonym.

4 Prognozowanie od ogółu do szczegółu

Kolejne badanie dotyczyło problemu prognozowania od ogółu do szczegółu (*top-down forecasting*), tj. problemu dzielenia prognoz wyższego rzędu na prognozy niższego

rzędu [25]. Prognoza wyższego rzędu może być prognozą zapotrzebowania na całą grupę produktów. Taka prognoza może być wartością wskazaną przez eksperta posiadającego wiedzę dziedzinową lub może to być prognoza pochodząca z innego modelu predykcyjnego. Prognoza niższego poziomu to prognoza zapotrzebowania na konkretny produkt. Uogólniając, zadanie prognozowania od ogółu do szczegółu sprowadza się do określenia wagi w , przez którą należy pomnożyć prognozę dla całej kategorii, aby uzyskać prognozę dla konkretnego produktu.

W badaniach przetestowano metodę naiwną, dostosowaną do własnych potrzeb metodę najbliższego sąsiada (KNN), parametryczny liniowy model mieszany (LMM) oraz metodę XGBoost. Badania zostały przeprowadzone na podstawie danych od firmy zajmującej się sprzedażą odzieży. Prognozy dla określonej kategorii produktów i prognozy dla poszczególnych produktów zostały nam dostarczone przez partnera biznesowego. Dane prognozy dotyczące konkretnych produktów zostały uznane za punkt odniesienia przy ocenie otrzymanych wyników. Badania przedstawione w tym rozdziale zostały również częściowo przedstawione w artykule [12].

Metoda naiwna obliczała wagę w na podstawie sprzedaży mającej miejsce dokładnie rok wcześniej.

Następnie wybrano metodę K-Nearest Neighbors (KNN) [16] jako rozszerzenie metody naiwnej. Algorytm KNN dokonał prognozy wagi w poprzez wyszukiwanie najbardziej podobnych obserwacji z całego zakresu danych historycznych. W pracy przedstawiono specjalne przygotowanie danych wejściowych. Ważna była zmiana wartości atrybutu *size* opisujące rozmiar produktu. Dany atrybut miał różną konwencję nazw dla różnych produktów. Aby obliczyć odległości między różnymi obserwacjami (różnymi produktami) przy użyciu atrybutu rozmiaru, historyczny zbiór danych został podzielony według danego atrybutu – oddzielny podzbiór powstał dla każdej kategorii produktów, które miały różne rozmiary odzieży. Następnie dla każdego z rozmiarów w zestawie określono wartości percentyla w tych podzbiorach i dana wartość stawała się nową wartością atrybutu. Atrybut nominalny *product type* (model produktu) również został przekształcony. W celu uzyskania odległości między różnymi typami produktów zaproponowano wykorzystanie *product types dissimilarity table* – tabelę niepodobieństwa modeli produktów.

Innym testowanym modelem były liniowe modele mieszane (LMM), które są rozszerzeniem prostych modeli regresji liniowej i mogą być stosowane do danych o strukturze hierarchicznej, którą obserwuje się w modzie. Modele te uwzględniają efekty stałe i losowe. W tym podejściu oszacowano model dla każdej grupy produktów

z losowymi efektami zdefiniowanymi przez model produktu, rozmiar i kolor. Liniowe modele mieszane wykorzystywały również *product types dissimilarity table* i zmiany wartości atrybutów podobne do podejścia KNN.

Ostatnią badaną metodą była metoda XGBoost. Zadaniem modelu było również prognozowanie wagi, przez którą należy pomnożyć prognozę kategorii, aby uzyskać prognozę dla konkretnego produktu. Model prognozowania wykorzystywał te same cechy, co modele KNN i LMM.

Testy statystyczne zostały wykorzystane w celu oceny różnic pomiędzy błędami predykcji pomiędzy testowanymi metodami. Najpierw przeprowadzono nieparametryczny test Friedmana [8], który służy do porównywania ze sobą wyników więcej niż dwóch algorytmów. Test ten sprawdza hipotezę zerową, mówiącą o tym, że wszystkie algorytmy są takie same. Hipotezę zerową można było odrzucić, więc przeprowadzono analizę *post-hoc* przy użyciu testu Nemenyi [19], a wyniki przedstawiono za pomocą wykresów różnic krytycznych (wykresy CD) [7]. Diagram ten pokazuje grupy algorytmów, takie, że algorytmy znajdujące się w różnych grupach są od siebie statystycznie różne na podstawie wyników procedury Nemenyi.

Ranking metod wskazywał, że biorąc pod uwagę wszystkie produkty, najlepsze wyniki dała metoda LMM bez optymalizacji. Wykres CD nie wykazał statystycznie istotnej różnicy między średnimi rangami metody LMM bez optymalizacji i wyników bazowych (dostarczonych przez partnera biznesowego). Test Nemenyi jest jednak uważany za konserwatywny, więc przeprowadzono również test Wilcoxona z korektą p-wartości dla porównań wielokrotnych [22]. Test ten wykazał znaczące różnice na poziomie 0,05, co oznacza, że na podstawie testu Wilcoxona można stwierdzić, że zaproponowana metoda LMM bez optymalizacji dała znacznie lepsze wyniki niż prognozy bazowe.

Przeprowadzone badania proponują wykorzystanie metod uczenia maszynowego do realizacji zadania prognozowania od ogółu do szczegółu. W kontekście prognozowania dla branży modowej warto zwrócić uwagę na zastosowaną tu metodę przygotowania danych w taki sposób, aby można je było wykorzystać w metodach wykorzystujących miarę odległości. Atrybut “rozmiar” został zmieniony na wartość liczbową przy użyciu wartości percentyla, które odpowiadają konkretnemu rozmiarowi w zestawie danych historycznych. Dla atrybutu “model produktu”, który służył jako identyfikator i nie miał dodatkowego znaczenia, zaproponowano tabelę niepodobieństwa modeli produktów w celu uzyskania odległości między analizowanymi produktami. Taka

tabela może być określona na podstawie danych (na podstawie informacji o bardziej ogólnych kategoriach produktów), ale może być również określona przez eksperta.

5 Włączenie dodatkowych szeregów czasowych związanych z prognozowaniem zapotrzebowania

W kolejnym rozdziale przedstawiono możliwości i zalety wykorzystania dodatkowych szeregów czasowych w celu poprawy wyników prognozowania.

5.1 Włączanie danych z podobnych lokalizacji

W pierwszej kolejności pokazano możliwość wykorzystania analogicznych szeregów czasowych z innych lokalizacji. Przedstawiono różne możliwości grupowania i dodawania informacji z podobnych szeregów czasowych. Przetestowano następujące podejścia:

1. Prognoza na podstawie jednej wartości – Modele prognostyczne zostały utworzone wyłącznie na podstawie danych historycznych z rozważanego szeregu czasowego. Nie wykorzystano żadnych dodatkowych szeregów czasowych. Wyniki tych eksperymentów były punktami odniesienia. Metody, które zostały wzięte pod uwagę to: średnia ruchoma, ARIMA [3], Prophet [23], XGBoost i sieć Long Short-Term Memory (LSTM) [14].
2. Prognoza po umieszczeniu wszystkich zestawów danych w jednej grupie – W tych eksperymentach w modelu prognostycznym wykorzystano dane ze wszystkich lokalizacji. Szeregi czasowe nie zostały pogrupowane w żaden szczegółowy sposób. Rozważono metodę LSTM i XGBoost w dwóch wersjach.
3. Prognoza na podstawie podobieństwa rekordów – W tym eksperymencie zastosowano typową metodę KNN dla regresji.
4. Prognoza na podstawie grupowania szeregów czasowych – Zbiory danych zostały pogrupowane według charakterystyki ich znormalizowanych szeregów czasowych. Następnie na podstawie tych grup stworzono modele prognostyczne. Rozważono metodę LSTM i XGBoost w dwóch wersjach.

5. Prognoza oparta na podwójnym grupowaniu – W tym eksperymencie dane ze wszystkich szeregów czasowych zostały znormalizowane i pogrupowane według charakterystyki szeregu czasowego. Następnie dla każdej grupy dla każdego znormalizowanego zbioru danych dodano nową kolumnę logiczną określającą przyporządkowanie do wyznaczonej grupy. Następnie wszystkie zbiory danych zostały połączone w jeden zbiór. Obliczono optymalną liczbę klastrów dla algorytmu *k*-średnich (*k-means*), a następnie przeprowadzono grupowanie metodą *k-means*. Dla każdej grupy utworzono model XGBoost.

Każde z tych podejść zostało szczegółowo opisane w rozprawie doktorskiej.

W badaniu wykorzystano dwa zbiory danych. Pierwszy zestaw danych opisywał sprzedaż paliwa na wielu stacjach benzynowych. Dane te zostały wykorzystane w badaniach przedstawionych w artykule [9]. Drugi zestaw danych opisywał wykorzystanie gazu LPG w zbiornikach zlokalizowanych w pobliżu budynków [17]. Te dwie grupy zbiorów danych zawierały tylko kilka dodatkowych atrybutów warunkowych, więc inżynieria cech była kluczowym krokiem.

Ponieważ wiele podejść zostało przetestowanych na wielu zestawach danych, dlatego wykonano testy za pomocą testu Friedmana i korekty Imana Davenporta. Następnie przeprowadzono analizę *post-hoc* – przeprowadzono test Nemenyi i test Wilcoxona z korektą p-wartości dla porównań wielokrotnych. Biorąc pod uwagę wykresy CD prezentujące wyniki testu Nemenyi oraz analizując wyniki testów Wilcoxona, najlepsze wyniki uzyskiwano zawsze dla metod wykorzystujących dane z wielu lokalizacji. W obu przypadkach, w najlepszej grupie wskazanej przez wykresy CD, najlepsze podejścia obejmowały te same 3 procedury. Pierwszą z nich była zaproponowana metoda podwójnego grupowania. Druga metoda polegała na połączeniu wszystkich zbiorów w jeden zbiór danych treningowych, który został wykorzystany do trenowania modelu prognozującego znormalizowaną wartość atrybutu decyzyjnego. Trzecie podejście wykorzystywało grupowanie szeregów czasowych. W tym podejściu, po utworzeniu grup na podstawie charakterystyki szeregów czasowych, dla każdej grupy utworzono oddzielny zestaw treningowy. Następnie dane opisujące szeregi czasowe przydzielone do jednej grupy zostały połączone (po wcześniejszej normalizacji), a następnie na podstawie tych zestawów wytrenowano modele predykcyjne, które przewidywały znormalizowaną wartość w ramach tych grup. W przypadku prognozy zużycia gazu LPG również metoda ARIMA wypadła bardzo dobrze. Biorąc pod uwagę podane wyniki, można stwierdzić, że dla badanych przypadków wprowadzenie danych dotyczących podobnego problemu z innych lokalizacji znacznie poprawiło

wyniki predykcji, a podane podejście może być warte rozważenia w podobnych sytuacjach.

5.2 Użycie danych z tej samej lokalizacji

Inną kwestią podobną do problemu włączania danych z podobnych lokalizacji jest włączanie danych z tej samej lokalizacji. Skupiono się na dodaniu danych z różnych szeregów czasowych, które mogą być mierzone w tej samej lokalizacji, co przewidywany szereg czasowy, w celu uzyskania lepszych prognoz. Kwestia ta została zilustrowana na przykładzie prognozowania zużycia energii w budynkach. Prezentowane badania zostały przeprowadzone w kontekście projektu, który dotyczył opracowania modelu cyfrowego bliźniaczego budynku. Dzięki temu możliwe było wykorzystanie informacji o zużyciu energii przez wybrane urządzenia gospodarstwa domowego. Dane te, wraz z informacjami o całkowitym zużyciu energii w danej lokalizacji i danymi pogodowymi, zostały wykorzystane w podejściach prognostycznych. Podejście to jest unikalne w porównaniu z innymi pracami. Badania zostały również zaprezentowane w artykule [13].

Zbadano pięć różnych lokalizacji, wyposażonych w urządzenia z licznikami zużycia energii. Monitorowano również zużycie energii w całej lokalizacji. Model bazowy został stworzony przy użyciu podejścia naiwnego. Model ten przewidywał wartość zaobserwowaną w danej lokalizacji dokładnie tydzień wcześniej. Następnie stworzono cztery modele oparte na regresji liniowej. Metody LSTM i Prophet zostały zastosowane jako bardziej zaawansowane metody do zadań prognozowania szeregów czasowych. Przeprowadzono testy również dla różnych zestawów danych. Rozpatrzono następujące scenariusze:

- Użycie tylko informacji o całkowitym zużyciu energii.
- Wykorzystanie informacji o całkowitym zużyciu energii i pogodzie.
- Wykorzystanie informacji o całkowitym zużyciu energii i zużyciu energii przez 10 najbardziej energożernych urządzeń.
- Wykorzystanie informacji o całkowitym zużyciu energii, pogodzie i zużyciu energii przez 10 najbardziej energożernych urządzeń.

Celem badań było uzyskanie błędu predykcji zużycia energii mniejszego niż 25% i udało się to osiągnąć dla 4 z 5 lokalizacji. Dla tych czterech lokalizacji najlepsze wyniki

uzyskano w eksperymencie, w którym zastosowano metodę Prophet z informacjami o całkowitym zużyciu energii i zużyciu energii przez 10 urządzeń.

W rozprawie przedstawiono również metodę analizy błędów modelu prognostycznego. W tym celu utworzono modele klasyfikacji dla każdej lokalizacji. Dla każdej lokalizacji i jej modelu Prophet prognoza była oznaczana jako *Poprawna*, gdy błąd predykcji był mniejszy niż 25% i *Niepoprawna*, gdy błąd predykcji był równy lub większy niż 25%, lub gdy model nie zwrócił predykcji. Te etykiety zostały potraktowane jako dwie wartości decyzyjne w zadaniu klasyfikacji binarnej. Następnie utworzono modele drzew decyzyjnych przy użyciu zaproponowanych atrybutów warunkowych. Cechy te opisywały wartości zużycia energii w porównaniu do poprzednich wartości historycznych oraz zmiany trendu zużycia energii. Szczegółowy opis procesu generowania cech do tworzenia modelu drzewa decyzyjnego został zawarty w rozprawie. Modele decyzyjne pokazały, które urządzenia, a dokładniej ich charakterystyka zużycia energii, miały największy wpływ na wyniki prognoz. Wreszcie, stworzone modele drzew decyzyjnych zostały wykorzystane do ograniczenia liczby monitorowanych urządzeń. W większości przypadków usunięcie z modelu Prophet informacji o urządzeniach, które pojawiły się w drzewie, dało wyniki bardzo zbliżone do sytuacji, w której monitorowano wszystkie 10 urządzeń.

Podsumowując badania przedstawione w danym rozdziale rozprawy, przeanalizowano potencjał włączenia informacji z innych szeregów czasowych, pochodzących z tej samej lub innej lokalizacji. We wszystkich badanych przypadkach dodatkowe informacje spowodowały poprawę prognoz.

6 Zwroty

Ostatnim rozważanym zagadnieniem był problem prognozowania prawdopodobieństwa zwrotu konkretnego produktu z konkretnej transakcji zakupu. Sprawdzono możliwość wykorzystania do tego zadania metod uczenia maszynowego. Przegląd literatury wykazał, że temat prognozowania zwrotów przy użyciu metod uczenia maszynowego nie był szeroko analizowany i może stanowić potencjalną lukę badawczą.

Celem badania było stworzenie modelu predykcyjnego na podstawie danych opisujących transakcje zakupowe. Dla każdego produktu w każdej transakcji zamierzano dodać binarny atrybut decyzyjny, który wskazywałby, czy produkt został zwrócony. Jednak po przeanalizowaniu dostępnych danych okazało się, że w dostępnej bazie danych sprzedawcy nie było żadnych informacji o tym, która transakcja zwrotu była

powiązana z którą transakcją zakupu. Z tego powodu nie było możliwości, aby wprost otrzymać informację o tym, czy zakupiony produkt został zwrócony. W związku z tym zaproponowano własny algorytm łączenia transakcji zwrotu z odpowiadającą mu transakcją zakupu.

Algorytm składa się z dwóch kroków. W pierwszym kroku, dla rozważanego zwrotu, podejmowana jest próba znalezienia transakcji zakupu, która ma wysokie prawdopodobieństwo bycia odpowiednią transakcją zakupu. Będzie to transakcja zakupu, która miała miejsce w ograniczonym czasie przed zwrotem i zawiera wszystkie produkty, które zostały uwzględnione w zwrocie. Jeśli w bazie danych znajduje się tylko jedna taka transakcja, można powiedzieć, że odpowiadająca transakcja zakupu została znaleziona z dużą pewnością i algorytm zostaje zakończony. Jeśli istnieje wiele transakcji zakupu spełniających te kryteria, algorytm przechodzi do kroku drugiego. Jeśli żadna transakcja zakupu nie spełnia warunków z kroku pierwszego, algorytm również zostaje zakończony z informacją, że nie znaleziono pasującej transakcji zakupu dla rozważanego zwrotu.

W kroku drugim podejmowana jest próba wybrania pasującej transakcji zakupu spośród wcześniej zawężonej grupy. Wybór pasującej transakcji zakupu opiera się na podobieństwie kanału transakcji (zakup i zwrot miały miejsce online lub w sklepie), na liczbie podobnych produktów w transakcji zakupu (wiele podobnych przedmiotów może sugerować, że klient kupił produkty z zamiarem zwrotu niektórych z nich) oraz na podstawie wartości koszyka. Jeśli po tych wszystkich krokach nadal istnieje więcej niż jedna pasująca transakcja zakupu, ostateczny wybór odpowiadającej transakcji zakupu jest dokonywany losowo, ale już w mocno ograniczonej grupie. Schematy blokowe dla zaproponowanego algorytmu zostały przedstawione w rozprawie.

Ze względu na brak informacji o realnych powiązaniach między transakcjami, nie jest możliwe jednoznaczne zweryfikowanie zaproponowanego algorytmu. Można jednak zauważyć, że wiele transakcji zwrotnych zostało powiązanych z dużą pewnością z jedną pasującą transakcją zakupu (40,6%). W przypadku pozostałych transakcji tylko dla małej grupy z nich nie znalazło żadnej pasującej transakcji zakupu (2,9%), a niewielka liczba zwrotów została powiązana przy użyciu losowego wyboru (1,1%). Biorąc to pod uwagę, proponowany algorytm wydaje się być dobrym rozwiązaniem w celu łączenia transakcji zwrotów z transakcjami zakupu w przypadku, gdy nie ma takich powiązań zapisanych w bazie danych. Te powiązania umożliwiają utworzenie atrybutu decyzyjnego w zbiorze treningowym i utworzenie innych atrybutów warunkowych opisujących transakcje zakupu.

W prezentowanych badaniach wzięto pod uwagę trzy algorytmy, które zostały wykorzystane do utworzenia modeli prognozujących, czy produkt zostanie zwrócony: regresja logistyczna [15], Random Forest [4] i XGBoost. Wszystkie algorytmy zostały przetestowane w 4 konfiguracjach.

W pierwszym podejściu wykorzystano tylko oryginalne atrybuty opisujące każdy produkt. Obejmowały one kolor, rozmiar, kategorię produktu, cenę. Zbiór danych opisywał również datę sprzedaży produktu, w tym informacje o świętach i wydarzeniach specjalnych.

Drugie podejście stanowiło rozszerzenie pierwszego. Przeprowadzono inżynierię cech i dodano atrybuty opisujące koszyk i inne zakupione produkty. Opiera się to na wiedzy dziedzinowej, wnioskach z przeglądu literatury i naszych własnych obserwacjach danych – gdy kupowanych jest wiele sztuk bardzo podobnych produktów lub tych samych tylko w różnych rozmiarach/kolorach, prawdopodobieństwo zwrotu z danej transakcji wzrasta. Nowymi atrybutami były na przykład liczba produktów w koszyku, wartość koszyka, liczba produktów w koszyku w tym samym kolorze, ale różnych rozmiarach oraz wskaźnik popularności produktu.

W trzecim podejściu przeprowadzono dostrajanie parametrów dla każdego modelu. Zoptymalizowano również wartość odcięcia wartości atrybutu decyzyjnego. W tym podejściu wykorzystano wszystkie wcześniej utworzone atrybuty.

Ostatnie podejście obejmowało balansowanie klas. Zastosowano technikę *Synthetic Minority Over-sampling Technique* (SMOTE). Podejście to obejmowało również wszystkie poprzednie modyfikacje.

Najlepszy wynik uzyskano dla modelu XGBoost dla podejścia numer 3 (po inżynierii cech i dostrojeniu parametrów) z wartością AUC równą 0,806. Uzyskany wynik jest zadowalający – inżynieria cech i dostrajanie parametrów dały dobre wyniki. Można stwierdzić, że dodanie cech opisujących zawartość koszyka poprawiło wyniki dla każdej badanej metody. Cechy te można było dodać dzięki stworzonemu algorytmowi łączącemu transakcję zakupu z transakcją sprzedaży. Pomimo braku rzeczywistych informacji na temat powiązań między transakcjami, możliwe było uzyskanie cech, które poprawiły model predykcyjny. Może to wskazywać na przydatność proponowanego algorytmu.

7 Podsumowanie

W pracy przedstawiono autorską taksonomię zagadnień związanych z prognozowaniem zapotrzebowania. Każde zadanie zostało opisane z uwzględnieniem specyficznych problemów, które mogą być związane z każdym z nich. Taksonomia ta może pokazać szerszy obraz zagadnień związanych z prognozowaniem zapotrzebowania. Następnie skupiono się na wybranych zagadnieniach i rozważono, w jaki sposób można zastosować metody uczenia maszynowego do rozwiązywania danych problemów. Praca doktorska nie koncentruje się na typowym zagadnieniu prognozowania szeregów czasowych oraz porównaniu wszystkich znanych algorytmów prognozowania szeregów czasowych. Zamiast tego skupiono się na mniej znanych kwestiach związanych z prognozowaniem zapotrzebowania.

Wkład rozprawy doktorskiej jest następujący:

- Zaproponowano taksonomię zagadnień związanych z prognozowaniem zapotrzebowania, pozwalającą na ustrukturyzowanie tematu i zapewniającą szerszą perspektywę niż tylko skupienie się na prognozowaniu szeregów czasowych.
- Zaproponowano sposób prognozowania efektu promocji przy użyciu sześciu wskaźników i modeli uczenia maszynowego. Podejście to zostało następnie wykorzystane w symulatorze wrażliwości cenowej i systemie rekomendacji promocji.
- Zaprezentowano oryginalny algorytm indukcji przeżyciowych reguł akcji w celu wskazania zmian, które wpływają na skuteczność promocji.
- Przedstawiono zastosowanie metod uczenia maszynowego do prognozowania od ogółu do szczegółu w branży odzieżowej. Zaproponowano metodę przygotowania cech przy użyciu *tabeli niepodobieństwa modeli produktów* i konwersji różnych rozmiarów produktów na wartość liczbową.
- Przeprowadzono badania, które wskazują na znaczenie wykorzystania dodatkowych szeregów czasowych do prognozowania zapotrzebowania dla rozważanych przypadków, biorąc pod uwagę szeregi czasowe z podobnych lokalizacji i tej samej lokalizacji.
- Wykorzystując informacje z modelu cyfrowego bliźniaczego budynku, do stworzenia modelu predykcyjnego zużycia energii wykorzystano dodatkowe szeregi

czasowe. Nasze badania wykazały, że monitorowanie zużycia energii nie tylko dla całej lokalizacji, ale także dla wybranych urządzeń, może znacznie poprawić zdolność prognozowania zużycia energii.

- Zaproponowano algorytm łączenia transakcji zwrotu z transakcją zakupu. Przeprowadzono również badania prognozowania zwrotu określonego produktu z określonej transakcji zakupu, osiągając dobre wyniki prognozowania.
- Wkład pracy obejmuje również zaprezentowaną inżynierię cech oraz prezentację atrybutów, które warto wykorzystać w konkretnych przypadkach prognozowania zapotrzebowania.
- W rozprawie przedstawiono mniej znane zagadnienia związane z prognozowaniem zapotrzebowania.

Przeprowadzone badania potwierdziły również postawioną tezę. Zaproponowana taksonomia problemów prognozowania popytu pozwoliła na podejście do nich w dopasowany sposób z wykorzystaniem wiedzy dziedzinowej i metod uczenia maszynowego. Wiedza dziedzinowa została wprowadzona do badań poprzez podjęcie decyzji o zakresie danych wykorzystywanych w modelu predykcyjnym, w kontekście tego, co ma być prognozowane i dla jakich horyzontów prognozy. We wszystkich badaniach wiedza dziedzinowa przejawiała się również w kontekście inżynierii cech.

Wszystkie badania zostały przeprowadzone na rzeczywistych danych. Większość z nich pochodziła od dużych firm z wieloma oddziałami. Dane często obejmowały kilka lat, pochodziły z wielu lokalizacji i dotyczyły wielu produktów. Rozważane badania były podyktowane projektami badawczymi R&D, więc można stwierdzić, że są one istotne dla biznesu. Część badań przedstawionych w rozprawie była związana z następującymi projektami: “System wspomaganie decyzji i zarządzania wiedzą dla branży handlu detalicznego (DSS-RETAIL)” (POIR.01.01.01-00-0871/17-00) oraz “System SupplAI automatyzujący i optymalizujący planowanie łańcucha dostaw, wykorzystujący metody inteligencji obliczeniowej – w tym prognozowania sprzedaży – do analizy online bardzo dużych zbiorów danych pozwalający na ograniczenie strat i zwiększenie efektywności działania firm branży handlu detalicznego” (POIR.01.01.01-00-0500/21) – oba projekty zostały dofinansowane z Funduszy Europejskich i zrealizowane wspólnie z firmą 3Soft S.A. “System wspomaganie decyzji i zarządzania wiedzą operacyjną i procesową dla rynku dystrybutorów gazu LPG” współfinansowany przez Narodowe Centrum Badań i Rozwoju w ramach grantu

POIR.01.01.01-00-0104/17. Badania związane z prognozowaniem sprzedaży paliw na stacjach benzynowych zostały zrealizowane we współpracy z Działem FuelPrime w AIUT Sp. z o.o. i były częściowo wspierane przez Unię Europejską w ramach Europejskiego Funduszu Społecznego (grant POWR.03.05.00-00-Z305).

Pliki z danymi dla zadania krótkoterminowej prognozy zapotrzebowania na gaz LPG są dostępne na stronie <https://github.com/adaa-polsl/dss4lpg>. Witryna zawiera również oprogramowanie umożliwiające wykonywanie prostych analiz i generowanie raportu. Zbiory danych opisujące szeregi czasowe sprzedaży paliw zostały udostępnione pod adresem <http://adaa.polsl.pl/index.php/datasets-software/>.

Literatura

- [1] Badura, J., Hermansa, M., Kozielski, M., Sikora, M., Wróbel, Ł. Separate-and-conquer Survival Action Rule Learning. *Knowledge-Based Systems*, 2023 (Zgłoszono do czasopisma Knowledge-Based System. Artykuł po pierwszej rundzie recenzji. Stan na 22.05.2023).
- [2] Blachnik, M., Henzel, J. Estimating the Performance Indicators of Promotion Efficiency in FMCG Retail. Yang, H., Pasupa, K., Leung, A. C.-S., Kwok, J. T., Chan, J. H., King, I., redaktorzy, *Neural Information Processing*, strony 320–332. Springer International Publishing, 2020.
- [3] Box, G. E., Jenkins, G. M., Reinsel, G. C., Ljung, G. M. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [4] Breiman, L. Random Forests. *Machine Learning*, 45(1):5–32, 2001.
- [5] Chen, T., Guestrin, C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, wolumen KDD '16, strony 785–794. ACM, 2016.
- [6] Dardzinska, A. *Action rules mining*, wolumen 468. Springer, 2012.
- [7] Demšar, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.*, 7:1–30, 2006.
- [8] Friedman, M. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, 32(200):675–701, 1937.

- [9] Henzel, J., Bularz, J., Sikora, M. Impact of time series clustering on fuel sales prediction results. Ganzha, M., Maciaszek, L., Paprzycki, M., Ślęzak, D., redaktorzy, *Position and Communication Papers of the 16th Conference on Computer Science and Intelligence Systems*, wolumen 26 serii *Annals of Computer Science and Information Systems*, strona 13–21. PTI, 2021.
- [10] Henzel, J., Sikora, M. Gradient Boosting and Deep Learning Models Approach to Forecasting Promotions Efficiency in FMCG Retail. Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J. M., redaktorzy, *Artificial Intelligence and Soft Computing*, strony 336–345, Cham, 2020. Springer International Publishing.
- [11] Henzel, J., Sikora, M. Gradient Boosting Application in Forecasting of Performance Indicators Values for Measuring the Efficiency of Promotions in FMCG Retail. Agarwal, S., Barrell, D. N., Solanki, V. K., redaktorzy, *Proceedings of the 2020 Federated Conference on Computer Science and Information Systems*, wolumen 21 serii *Annals of Computer Science and Information Systems*, strony 59–68. IEEE, 2020.
- [12] Henzel, J., Wawrowski, Ł., Kubina, A., Sikora, M., Wróbel, Ł. Demand forecasting in the fashion business — an example of customized nearest neighbour and linear mixed model approaches. *Proceedings of the 17th Conference on Computer Science and Intelligence Systems (FedCSIS)*, strony 61–65, 2022.
- [13] Henzel, J., Wróbel, Ł., Fice, M., Sikora, M. Energy Consumption Forecasting for the Digital-Twin Model of the Building. *Energies*, 15(12), 2022.
- [14] Hochreiter, S., Schmidhuber, J. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [15] Hosmer, D. W., Lemeshow, S. *Applied Logistic Regression*. Wiley, wydanie 2nd, 2000.
- [16] Kordos, M., Strzempa, D., Blachnik, M. Do We Need Whatever More than k-NN? *LNCS*, 6113:414–421, 2010.
- [17] Kozielski, M., Henzel, J., Wróbel, Ł., Łaskarzewski, Z., Sikora, M. A Sensor Data-Driven Decision Support System for Liquefied Petroleum Gas Suppliers. *Applied Sciences*, 11(8), 2021.

- [18] Lawrence, M., Goodwin, P., O'Connor, M., Önköl, D. Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3):493–518, 2006.
- [19] Nemenyi, P. B. *Distribution-Free Multiple Comparisons*. Praca doktorska, Princeton University, 1963.
- [20] Sikora, M. Wybrane metody oceny i przycinania reguł decyzyjnych. *Studia Informatica*, 33(3B):5–331, 2012.
- [21] Stevenson, M. An Introduction to Survival Analysis. EpiCentre, IVABS, Massey University, 2007.
- [22] Stąpor, K., Ksieniewicz, P., García, S., Woźniak, M. How to design the fair experimental classifier evaluation. *Appl. Soft Comput.*, 104:107219, 2021.
- [23] Taylor, S. J., Letham, B. Forecasting at Scale. *The American Statistician*, 72(1):37–45, 2018.
- [24] Wróbel, Ł., Gudyś, A., Sikora, M. Learning rule sets from survival data. *BMC Bioinformatics*, 18(285):1–13, 2017.
- [25] Zotteri, G., Kalchschmidt, M., Caniato, F. The impact of aggregation level on forecasting performance. *International Journal of Production Economics*, 93-94:479–491, 2005.