

doi:10.15199/48.2024.08.10

Prognoza czasu wykonywania zadań przez pracowników floty pojazdów dla planowania pracy floty pojazdów elektrycznych pod kątem ich ładowania

Streszczenie. Artykuł przedstawia analizę statystyczną zmiennych wejściowych służących potencjalnie do stworzenia modeli prognoz czasu wykonywania zadań przez pracowników korzystających z floty pojazdów. Po wyborze różnych kombinacji zmiennych wejściowych wykonano prognozy za pomocą metod uczenia maszynowego, w tym sztucznych sieci neuronowych, lasów losowych i drzew decyzyjnych gradientowo wzmacnianych. Zakończenie artykułu stanowią wnioski z wykonanych badań oraz ocena przewidywanej użyteczności prognoz czasu wykonywania zadań w potencjalnej optymalizacji pracy floty pojazdów elektrycznych, w tym ich ładowania.

Abstract. The paper presents statistical analysis of input variables potentially useful for prediction of task execution time by vehicle fleet -using employees. After selection of inputs combinations machine learning was used for forecasts. Among others artificial neural networks, random forests and gradient boosted decision trees. Summary of the paper consists of conclusions, including the assessment of expected usefulness of predictions of task completion time for potential optimization of EV fleet work management including charging. (Forecast of task execution time by vehicle fleet employees for electric vehicle fleet work planning with regard to charging).

Słowa kluczowe: pojazdy elektryczne, prognozowanie, optymalizacja floty pojazdów, uczenie maszynowe
Keywords: electric vehicles, forecasting, vehicle fleet optimization, machine learning

Wstęp

Usługodawcy używający w swoich działalnościach floty pojazdów coraz częściej rozważają ich elektryfikację. Jest to związane zarówno z czynnikami ekonomicznymi jak i ograniczeniami w poruszaniu się aut spalinowych w miastach ale także z różnego rodzaju zachętami ze strony państwa. Zwykle operatorzy flot starają się maksymalizować użycie poszczególnych pojazdów, gdyż wówczas uzyskują najwyższą stopę zwrotu z zainwestowanego kapitału. Stosowane rozwiązania organizacyjne zarządzania flotami pojazdów spalinowych nie zawsze dają się łatwo przenieść na floty elektryczne. Jest to, rzecz jasna, związane z różnicami w sposobie użytkowania pojazdów elektrycznych i pojazdów spalinowych. Kluczowe są tu dwa elementy: zasięg pojazdu i czas ładowania akumulatorów. W niektórych przypadkach istotne okazują się dodatkowe wymagania dotyczące zwiększenia żywotności akumulatorów – np. okresowe balansowanie cel.

Ze względu na wspomniane elementy, zagadnie ustalenia optymalnego składu floty pojazdów elektrycznych staje się problemem bardzo skomplikowanym. Najczęściej nie wystarczy proste zastąpienie pojazdów spalinowych pojazdami elektrycznymi [1]. Zwykle należy także wprowadzić pewne zmiany w sposobie użytkowania floty pojazdów lub wprowadzać floty złożone z obu rodzajów pojazdów. Wyznaczanie rozwiązania optymalnego może wymagać wielokrotnego rozwiązywania skomplikowanych problemów optymalizacyjnych z rodziny Vehicle Routing Problem [2] [3] ze szczególnym uwzględnieniem problemów VRP z oknami czasowymi [2]. Określenie rozwiązania optymalnego tego rodzaju problemów nie jest możliwe bez dokładnych informacji definiujących to zadanie. Jedną z tych informacji jest przewidywany czas postoju pojazdów w poszczególnych punktach trasy – co w przypadku niniejszej pracy przekłada się na wymieniony w tytule czas wykonywania zadań przez pracowników korzystających z pojazdów. Oszacowanie tego czasu jest również niezbędne przy późniejszym operacyjnym wykorzystaniu floty pojazdów, której skład został wcześniej zoptymalizowany.

Dane wejściowe i wyjściowe

Zmienną prognozowaną (wyjściową procesu) stanowił czas wykonania zadania w zadanej lokalizacji przez brygadę techniczną korzystającą z pojazdu floty samochodów podlegającej optymalizacji czasu pracy. Każde zadanie stanowiło osobny przypadek zależny od różnych czynników, a problem pozbawiony był regularnych wyznaczników typowych dla prognozowania – tj. horyzontu prognozy, regularności odczytów w kwancie, itp., co stanowiło wyzwanie ze względu na brak możliwości zastosowania typowych zmiennych szeregu czasowego użytecznych dla klasycznej autoregresji. W celu usunięcia danych odstających zmienna wyjścia została zawężona do wartości mniejszych lub równych swojemu 99百分ylowi. Zebrane dane pochodziły z systemu zarządzania flotą stosowanego przez usługodawcę teletechnicznego z jednego z krajów europejskich. Zmienne wejścia i wyjścia zostały przedstawione w Tabeli 1. Przed wykonaniem badań wszystkie zmienne ciągłe zostały znormalizowane do zakresu 0-1 metodą min-max, a zmienne kategoryczne (BR, TC, TP, RD) zostały zapisane binarnie. Zmienne PU, NS, IWD, H potraktowano do obliczeń jako ciągłe, ze względu na to, że są to zmienne uporządkowane oraz ich wartości pośrednie mają logiczną interpretację.

Zmienne wejściowe stosowane w badaniach zostały podzielone na 8 kategorii:

1. zmienne kategoryczne związane z procesem,
2. lokalizacja zadania,
3. zmienne czasowe,
4. meta dane o koordynatach najbliższych punktów prognoz pogody względem miejsca wykonania zadania,
5. prognozy pogody,
6. gęstość zaludnienia docelowa,
7. zgrubna estymata czasu wykonania zadania opracowana przez właściciela floty,
8. średni czas dla zadania danego typu i rodzaju.

Takie ujęcie problemu umożliwiło uwzględnienie w prognozach zarówno aspektów technicznych (typ zadania, umiejętności brygad, etc., z kat. 1) jak i warunków lokalnych (trudności wykonywania prac w terenie ze względu na warunki atmosferyczne czy potencjalna trudność dostępu do infrastruktury wyrażona pośrednio gęstością zaludnienia).

Tabela 1. Zmienne wejścia i wyjścia stosowane w badaniach

Zmienna	OPIS	Kategoria zmiennej
BR	Numer brygady	1
TC	Numer technika	1
TP	Typ zadania	1
RD	Rodzaj zadania	1
PU	Wymagany poziom umiejętności do wykonania zadania	1
NS	Liczba umiejętności potrzebna do wykonania danego zadania	1
DG	Długość geograficzna	2
SG	Szerokość geograficzna	2
IWD	Dzień tygodnia	3
H	Godzina zegarowa	3
WID1	Numer węzła siatki meteo 1	4
WID2	Numer węzła siatki meteo 2	4
WR1	Rząd siatki meteo 1	4
WC1	Kolumna siatki meteo 1	4
WR2	Rząd siatki meteo 2	4
WC2	Kolumna siatki meteo 2	4
MN1	Numer oczka siatki 1	4
MN2	Numer oczka siatki 2	4
T	Temperatura powietrza	5
a10	Kierunek wiatru na 10 m	5
a30	Kierunek wiatru na 30 m	5
lc	Ilość niskich chmur	5
pr	Ilość opadów deszczu	5
sn	Ilość opadów śniegu	5
tc	Zachmurzenie całkowite	5
v10	Prędkość wiatru na 10 m	5
v30	Prędkość wiatru na 30 m	5
pd	Gęstość zaludnienia	6
ETD	zgrubny szacunek	7
ave	Średni czas danego typu i rodzaju zadania	8
RTD	Rzeczywisty czas wykonania zadania	Zmienna wyjścia

Charakterystyka statystyczna zmiennych wejściowych

Zebrane dane poddano analizie statystycznej i seriom oględzin eksperckich celem interpretacji niejasności. Na szczególną uwagę zasługuje zmienna wejścia z kat. 7, będąca osądem eksperckim właściciela floty na temat tego ile powinno trwać zadanie określonego typu. Zmienną tą scharakteryzowano pod względem statystycznym celem porównania ze zmienną prognozowaną. Porównanie to zawarto w Tabeli 2. Wyniki jednoznacznie wskazują na rozbieżność zmiennych, co świadczy o trudności problemu prognostycznego, wynikającej najprawdopodobniej z czynników losowych związanych z pracą floty (interakcji z klientem, awarii itp.) których nie da się ekspercko określić w estymacji zgrubnej. Tym samym, zwiększa się potencjał stosowania metod uczenia maszynowego, które są w stanie lepiej oddać takie czynniki. Całościowo, po oczyszczeniu danych zebrano ok. 8000 rekordów. Każdy rekord odpowiadał jednemu zadaniu z jego czasem trwania.

Tabela 2 Charakterystyka statystyczna rzeczywistego czasu wykonywania zadań (RTD) oraz czasu estymaty (ETD)

Miara	ETD [j.w]	RTD [j.w]
średnia	0,396	0,268
Odch. Std.	0,058	0,180
min	0,000	0,000
p25	0,286	0,131
p50	0,375	0,240
p75	0,500	0,370
max	1,000	1,000

Dla zmiennych wejścia obliczono metryki powiązania ze zmienną wyjścia. Dla zmiennych ciągłych obliczono korelację r-Pearsona (R), korelację τ -Kendalla(τ),

informację wzajemną (mi), oraz współczynniki p dla korelacji (R_p , τ_p). Wspomniane metryki wyrażono formułami 1 do 4 [4-8]. Wyniki przedstawiono w Tabeli 3. Zastosowanie różnych metryk umożliwiło analizę odporną na niespełnienie któregoś z założeń dla nich przyjmowanych.

$$(1) R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{(\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2})(\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2})}$$

$$(2) mi = - \sum_{i=1}^N P(x_i) \log P(x_i) - \sum_{i=1}^N \sum_{j=1}^M P(x_i, y_j) \log P(x_i \vee y_j)$$

$$(3) \tau = \frac{\sum_{i \neq j=1}^N \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j)}{N(N-1)}$$

$$(4) pbs = \left(\frac{\bar{Y}_1 - \bar{Y}_0}{\bar{s}} \sqrt{\frac{N_1 N_0}{N(N-1)}} \right)$$

gdzie: x_i, y_i – wartości chwilowe dwóch zmiennych korelowanych, \bar{x}, \bar{y} – wartości średnie zmiennych korelowanych,

$P(x_i)$ – prawdopodobieństwo wystąpienia x_i , $P(x_i, y_j)$ – prawdopodobieństwo łączne wystąpienia x_i, y_j , $P(x_i \vee y_j)$ – prawdopodobieństwo warunkowe wystąpienia x_i pod warunkiem wystąpienia y_j , N, M – liczba próbek dla zmiennych x_i, y_j , \bar{Y}_0, \bar{Y}_1 – średnie grupowe zmiennej kategorycznej zapisanej jako odpowiednio 0 i 1, $N_0, (N)_1$ – liczba próbek zmiennej zapisanych jako 0 i 1, \bar{s} – odchylenie standardowe zmiennej ciągłej

Tabela 3 Wybrane metryki powiązania między zmiennymi ciągłymi, a rzeczywistym czasem wykonania zadania

Zmienna	R	R_p	τ	τ_p	mi
DG	-0,02539	0,02243	-0,00744	0,31541	0,04038
SG	0,07260	6,3E-11	0,05120	5,0E-12	0,02826
PU	-0,05378	1,3E-06	-0,04625	9,8E-08	0,02157
NS	0,13975	1,5E-36	0,05588	8,5E-12	0,03680
IWD	-0,02295	0,039056	-0,01162	0,14910	0,00614
H	-0,03310	0,002908	-0,02688	0,00056	0,00511
WID1	0,06887	5,6E-10	0,05333	7,8E-13	0,03058
WID2	0,06863	6,5E-10	0,05261	1,6E-12	0,03047
WR1	0,06896	5,3E-10	0,05484	4,7E-13	0,01955
WC1	-0,03143	0,00471	-0,01436	0,05538	0,03132
WR2	0,06871	6,1E-10	0,05493	5,3E-13	0,02750
WC2	-0,03154	0,00456	-0,01514	0,04342	0,02418
MN1	0,06139	3,3E-08	0,03883	3,4E-07	0,03097
MN2	0,05987	7,1E-08	0,03741	9,1E-07	0,03122
T	-0,02702	0,01509	-0,01644	0,02667	0
a10	0,01704	0,12556	0,00473	0,52356	0
a30	0,01588	0,15328	0,00589	0,42686	0,01064
lc	0,00142	0,89822	0,00208	0,78040	0,00198
pr	0,00270	0,80809	-0,00318	0,68419	0
sn	-0,00244	0,82658	-0,00071	0,93577	0,00222
tc	0,01630	0,14273	0,01154	0,12084	0
v10	-0,01791	0,10728	-0,01460	0,04904	0
v30	-0,00453	0,68387	-0,00685	0,35536	0
pd	-0,08573	1,1E-14	-0,05081	7,5E-12	0,00801
ave	0,28730	5,5E-152	0,19748	1,8E-125	0,06975

W Tabeli 3 zaznaczono sześć wartości o największym module dla korelacji i informacji wzajemnej oraz współczynniki p większe niż próg istotności statystycznej 0,05. Dla zmiennych ciągłych najbardziej znacząca okazała się średnia dla czasu wykonania zadania. Analizując pozostałe wyniki z Tabeli 3 można zauważyć, że żadna z wyznaczonych korelacji nie uzyskała dużych wartości. Świadczy to o cząstkowym wpływie zmiennych na analizowany proces, co dodatkowo uwydatnia trudność analizowanego problemu. Największą korelacją dla współczynników R charakteryzowała się liczba umiejętności

wymagana do wykonania zadania (NS). Drugą co do wielkości modułu tej korelacji była gęstość zaludnienia. W dalszej kolejności korelowały zmienne związane z położeniem lokalizacji zadania. Zarówno korelacja tau jak i R okazały się nieistotne statystycznie dla zmiennych pogodowych. Ponadto, korelacja tau okazała się nieistotna dla długości geograficznej i jednej ze współrzędnych siatki prognoz meteorolog. odpowiadającej długości geograficznej (WC1). Dla korelacji tau bardziej znaczące okazały się meta informacje o współrzędnych lokalizacji wykonywania zadania niż na temat jawnie podanych długości i szerokości geograficznych. Wartości informacji wzajemnej wskazywały dla odmiany, że długość geograficzna podawana pośrednio (WC1) oraz bezpośrednio (DG) mają jednak wpływ na analizowany proces. Dla *mi*, mniej powiązane okazały się zmienne odpowiadające szerokości geograficznej (WR1, SG), zaś bardziej informacje zgrubne o obszarze na którym odbywać się będzie zadanie (MN1, MN2). Jednocześnie informacja wzajemna potwierdziła nieznaczny wpływ niektórych zmiennych pogodowych na analizowany proces. Całościowo, ze zmiennych w Tabeli 3, za najbardziej znaczące dla procesu zmienne można uznać czasy średnie wykonania danego typu i rodzaju zadania, koordynaty położenia zadania, liczbę i poziom umiejętności wymaganych do wykonania zadania oraz gęstość zaludnienia w obszarze docelowym.

Dla zmiennych kategoriowych po zakodowaniu ich w formie binarnej obliczono ich współczynnik korelacji punktowo-dwuseryjnej (*pbs*) ze zmienną prognozowaną oraz informację wzajemną. Wyniki przedstawiono w Tabeli 4. Ze względu na licznosc zmiennych kategoriowych w kodowaniu binarnym (ponad 600 zmiennych, gdzie każda to kategoria zdekomponowana na występujące unikatowe wartości), w tabeli 4 pokazano wyniki zgodnie z dwoma miarami. Wyniki posortowane malejąco po module *pbs* należały do kategorii A, a po informacji wzajemnej do kategorii B.

Tabela 4 Najwyższe wartości współczynników korelacji punktowo-dwuseryjnej oraz informacji wzajemnej

Zmienna	<i>pbs</i>	<i>pbs_p</i>	<i>mi</i>	Kategoria
TP2	0,16824	1,7E-52	0,00065	A
RD5	0,16785	3,1E-52	0,00368	A
RD8	0,15539	6,2E-45	0,00012	A
RD3	-0,14572	1,1E-39	0,00200	A
TP3	-0,13555	1,6E-34	0,00034	A
BR125	0,06205	2,3E-08	0,04126	B
TC22	0,06559	3,4E-09	0,03947	B
TC234	0,06205	2,3E-08	0,03329	B
TC113	0,08227	1,2E-13	0,02160	B
TC54	0,07046	2,1E-10	0,02008	B

Wśród przedstawionych w Tabeli 4 zmiennych, wszystkie okazały się statystycznie istotne ($\alpha = 0,05$). Jednocześnie, w danych posortowanych po module można było zaobserwować zmienne bardziej korelujące z analizowanym procesem niż zmienne ciągle, zarówno pozytywnie jak i negatywnie. W przypadku pięciu parametrów na największej wartości informacji wzajemnej (kat. B) *pbs* charakteryzowało się ok. 2 razy mniejszą wartością niż dla kategorii A, podczas gdy *mi* dla kategorii B było podobne rzędem wielkości do *mi* uzyskiwanego dla zmiennych ciągłych. Można było również zauważyć, że przeważnie informacja wzajemna za najbardziej powiązane z procesem wskazała numer technika (TC), a *pbs* rodzaj wykonywanego zadania (RD). Celem zróżnicowania występowania zmiennych w większej skali obliczono występowanie danej kategorii danych w górnych 10 % wyników o największej wartości modułu *pbs* i *mi*. Wyniki zostały przedstawione w Tabeli 5.

Tabela 5 Liczbą wystąpień zmiennych kategoriowych danego typu w 10 % wyników o największej wartości parametru kryterialnego

Kryterium	BR	TC	TP	RD
moduł <i>pbs</i>	25	26	4	9
<i>mi</i>	39	24	0	1

Główną różnicą, którą można było zaobserwować między metrykami powiązania, był wpływ danych o parametrach zadania na analizowany proces. Informacja wzajemna wyraźnie wskazała dominujący charakter danych o zespole i w dalszej kolejności techniku wykonującym zadanie. Dla *pbs* wpływ ludzki (BR, TC) w porównaniu do wpływu cech zadania (TP, RD) rozkładał się w proporcji 80 do 20 % Tym samym dla zmiennych kategoriowych za najważniejsze można było uznać informacje o tym kto realizował dane zadanie.

Prognozowanie czasu wykonania zadania

W oparciu o wykonaną wcześniej analizę statystyczną dokonano wyboru zmiennych wejściowych. Docelowo utworzono pięć różnych zestawów (kombinacji) danych wejściowych, przedstawionych w Tabeli 6. Ze względu na wyniki analizy w badaniach nie stosowano zmiennych pogodowych, a jedynie te, dla których miary powiązania ze zmienną prognozowaną były w czołówce wyników dla danej miary.

Tabela 6. Zestawy danych wejściowych

Zestaw	DG	SG	PU	NS	IWD	H	WID1	WID2	WR1	WC1	WR2	WC2	MN1	MN2	pd	BR	TC	TP	RD	ave
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1
3	1	0	0	1	0	0	0	0	1	0	0	1	1	0	1	1	0	0	1	1
4	1	0	0	1	0	0	1	1	0	1	0	0	1	1	0	1	1	1	1	1
5	1	1	1	1	0	0	1	1	0	1	0	0	1	1	1	1	1	1	1	1

W prognozach posłużono się trzema metodami popularnymi w literaturze [9-12]: sztuczną siecią neuronową MLP, lasami losowymi (RF), oraz drzewami decyzyjnymi wzmocnionymi gradientowo (XGBT). Dla każdej metody prognoz wybór kombinacji hiperparametrów poprzedziły wstępne testy modeli mające na celu zgrubne określenie zakresów parametrów. Dla MLP rozważano w badaniach od 3 do 15 neuronów w pierwszej warstwie ukrytej, od 0 do 6 w drugiej, funkcję aktywacji warstwy ukrytej relu albo tanh, liczbę iteracji 200 albo 500 oraz licznik oczekiwania na spadek błędów (cierpliwość iteracyjną) od 10 do 50. Dla XGBT rozważano od 50 do 100 drzew, od 2 do 6 poziomów maksymalnej głębokości drzew, od 0.1 do 0.25 dla współczynnika uczenia. Dla lasów losowych rozważano od 50 do 400 drzew, od 3 do 30 poziomów maksymalnej głębokości drzew oraz od 1 do 10 cech branych pod uwagę podczas poszukiwania rozdziałów w drzewie.

W badaniach zastosowano walidację 5+1-krotną. Na początku wylosowano i odseparowano od reszty danych 20 % próbek jako dane testowe. Następnie, reszta danych była sześciokrotnie dzielona losowo na dane treningowe i walidacyjne w proporcji 80 do 20 %. Na pierwszych pięciu podziałach było wykonywane uczenie modeli, po czym wyniki na zbiorze walidacyjnym uśredniono celem wyboru kombinacji hiperparametrów modeli zapewniających najniższy wynik na tym zbiorze. Ostatecznie, zdolność generalizacji modeli określano na podziale szóstym. Procedurę stosowano dla każdego z zestawów zmiennych wejściowych. Jako miarę błędów zastosowano NRMSE [13] wyrażone formułą 5. Wyniki prognoz najlepsze dla danego zestawu danych wejściowych i typu modelu przedstawiono w Tabeli 7.

$$(5) \quad NRMSE = 100\% \cdot \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{y_i - \hat{y}_i}{C} \right)^2}$$

gdzie: y_i , \hat{y}_i –wartość rzeczywista i prognozowana zmiennej wyjścia, C – wartość maksymalna czasu wykonania zadania, N – liczba próbek,

Tabela 7. Wyniki prognoz

Metoda	Zestaw danych wejściowych	NRMSE Tren. [%]	NRMSE Walid. [%]	NRMSE Test [%]
XGBT	1	14,913	16,931	16,643
XGBT	2	14,968	16,849	16,485
XGBT	3	15,747	16,936	16,615
XGBT	4	16,261	16,848	16,709
XGBT	5	15,472	16,916	16,561
RF	1	12,410	16,960	16,647
RF	2	14,326	16,883	16,741
RF	3	14,935	17,274	16,829
RF	4	13,850	17,044	16,784
RF	5	13,210	16,949	16,679
MLP	1	15,899	16,686	16,450
MLP	2	15,723	16,970	16,591
MLP	3	15,832	17,268	16,877
MLP	4	15,907	16,926	16,639
MLP	5	15,856	16,761	16,420
Naiwna	zmienna ave	17,614	17,485	17,403

W Tabeli 7 zmienionym kolorem tła zaznaczono najlepsze na danych walidacyjnych modele danego typu. Pogrubieniem zaznaczono również model najlepszy z analizowanych. Dla XGBT najlepsze wyniki osiągnięto dla 100 drzew, głębokości maksymalnej drzew równej 2 i współczynnika uczenia równego 0.25. Dla lasów losowych najlepsze wyniki zapewniło 400 drzew o maksymalnej głębokości równej 30 i maksymalnej liczbie cech rozważanej przy podziale równej 10. Dla MLP najlepsze wyniki osiągnięto dla sieci z dwoma warstwami ukrytymi o 3 neuronach i funkcją aktywacji relu w każdej z nich, przy 500 iteracjach i cierpliwości iteracyjnej równej 50.

Podsumowanie i wnioski

Wśród 5 metod najlepszych, zarówno na danych walidacyjnych jak i testowych, najdokładniejsze okazały się modele MLP oraz XGBT. Względem metody naiwnej uzyskały one na danych testowych wyniki dokładniejsze, z poprawą sięgającą 5,65 % dla najlepszego modelu MLP. Jednocześnie nie stwierdzono wyraźnej przewagi w wynikach konkretnego zestawu danych wejściowych.

Zarówno MLP jak i XGBT uzyskiwały najlepsze wyniki na prostych strukturach informacyjnych i szybkim uczeniu (płytkie sieci z szybkim przerywaniem uczenia przy braku poprawy, płytkie drzewa, wysoki współczynnik uczenia). Dla odmiany lasy losowe wyraźnie preferowały głębokie, złożone struktury, z dużą liczbą głębokich drzew i cech rozpatrywanych przy podziale. Można to interpretować jako próbę rozwiązania przez metodę silnego czynnika losowego w danych. Dla MLP i XGBT ma to formę wyciągania jedynie najistotniejszych informacji z danych i dużych skoków w przestrzeni informacyjnej w celu ominięcia minimów lokalnych. Dla lasów losowych metoda próbuje zaś bardzo dokładnie podzielić dane na koszyki celem minimalizacji błędu. Całościowo potwierdza to trudność problemu prognostycznego przy zastanych danych i skłania ku wzbogaceniu systemu raportowania czasu wykonania zadań przez techników o składowe umożliwiające dalsze zmniejszenie błędu losowego – np. czasu rozmowy z klientem, czasu naprawy awarii, itp.

Przedstawione średnie błędy względne prognoz przekładają się na około 5 minutowy średni błąd oszacowania czasu wykonania zadania przez pracowników. Z pragmatycznego punktu widzenia, taki uchyb

oszacowania, w środowisku, w którym istnieje wiele interakcji klient-pracownik wydaje się akceptowalnym. Należy przy tym wskazać, że uchyb ten nie jest niestety jedynym źródłem niepewności w całym procesie biznesowym usługodawcy. Wśród innych należy wskazać błędy oszacowania: czasu dojazdu do klienta, czasu parkowania pojazdu, czasu pobierania materiałów z magazynu. Oznacza to, że zarówno metoda optymalizacji składu floty pojazdów jak i metoda optymalizacji przydziału zadań będą musiały uwzględniać te niepewności, aby wynik ich działania miał praktyczną wartość.

Artykuł bazował na wynikach projektu „System automatycznego przydzielania zadań w warunkach dużej zmienności wykorzystujący mechanizmy sztucznej inteligencji i optymalizacji geoprzestrzennej”, finansowanego ze środków NCBiR w ramach programu „Inteligentny Rozwój 2014-2020

Autorzy: dr hab. inż. Dariusz Baczyński, Politechnika Warszawska, Instytut Elektroenergetyki, ul. Koszykowa 75, 00-662 Warszawa, Polska, E-mail: dariusz.baczynski@pw.edu.pl; dr inż. Marcin Kopyt, Politechnika Warszawska, Instytut Elektroenergetyki, ul. Koszykowa 75, 00-662 Warszawa, Polska, marcin.kopyt@pw.edu.pl; Tomasz Gulczyński, Globema Sp. z o.o., ul. Wita Stwosza 22, 02-661 Warszawa, Polska, E-mail: tomasz.gulczynski@globema.pl

LITERATURA

- [1] Franceschetti, A., Honhon, D., Laporte, G., Van Woensel, T., Fransoo, J. C., Strategic fleet planning for city logistics, Transportation Research Part B: Methodological, Volume 95, 2017, 19-40, ISSN 0191-2615,
- [2] Hiermann, G., Puchinger, J., Ropke, S., Hartl, R. F. The Electric Fleet Size and Mix Vehicle Routing Problem with Time Windows and Recharging Stations, European Journal of Operational Research, Volume 252, Issue 3, 2016, 995-1018, ISSN 0377-2217.
- [3] Al-dal'ain, R., Celebi, D. Planning a mixed fleet of electric and conventional vehicles for urban freight with routing and replacement considerations, Sustainable Cities and Society, Volume 73, 2021, 103105, ISSN 2210-6707,
- [4] Gong, Huanhuan et al. A New Filter Feature Selection Algorithm for Classification Task by Ensembling Pearson Correlation Coefficient and Mutual Information. Engineering applications of artificial intelligence 131 (2024): 107865. Print.
- [5] Shi, Zhanhui et al. Identifying Reliability High-Correlated Gates of Logic Circuits With Pearson Correlation Coefficient. IEEE transactions on circuits and systems. II, Express briefs (2024): 1. Print.
- [6] Hajj Seyed Asadollah, Seyed Babak et al. An Intelligent Approach for Estimating Aeration Efficiency in Stepped Cascades: Optimized Support Vector Regression Models and Mutual Information Theory. Soft computing (Berlin, Germany) 26.24 (2022): 13969-13984. Print.
- [7] Xu, Weichao et al. A Comparative Analysis of Spearman's Rho and Kendall's Tau in Normal and Contaminated Normal Models. Signal processing 93.1 (2013): 261-276. Print. <https://doi.org/10.3390/en15249657>
- [8] Kornbrot, Diana. "Point Biserial Correlation." *Wiley StatsRef: Statistics Reference Online* (2014): Wiley StatsRef: Statistics Reference Online, 2014. Print.
- [9] Noorunnahar, Mst, Arman Hossain Chowdhury, and Farhana Arefeen Mila. A Tree Based eXtreme Gradient Boosting (XGBoost) Machine Learning Model to Forecast the Annual Rice Production in Bangladesh. PloS one 18.3 (2023): e0283452. Print.
- [10] Gono, Dylan Norbert, Herlina Napitupulu, and Firdaniza Firdaniza. Silver Price Forecasting Using Extreme Gradient Boosting (XGBoost) Method. Mathematics (Basel) 11.18 (2023): 3813. Print.
- [11] Moore, P. J., T. J. Lyons, and J. Gallacher. Random Forest Prediction of Alzheimer's Disease Using Pairwise Selection from Time Series Data. PloS one 14.2 (2019): e0211558. Print.
- [12] Gupta, Sonal, Deepankar Chakrabarty, and Rupesh Kumar. Predicting Indian Electricity Exchange-Traded Market Prices: SARIMA and MLP Approach. OPEC energy review 47.4 (2023): 271-286. Print.
- [13] Piotrowski, P.; Rutyna, I.; Baczyński, D.; Kopyt, M. Evaluation Metrics for Wind Power Forecasts: A Comprehensive Review and Statistical Analysis of Errors. Energies 2022, 15, 9657.