1. Artur WACHTARCZYK[1], 2. Janusz SOWIŃSKI[2]

Politechnika Częstochowska (1), Politechnika Częstochowska, Katedra Elektroenergetyki (2)
ORCID: 1. - ; 2. 0000-0002-1629-4977

doi:10.15199/48.2023.11.46

Analiza predykcyjna danych pomiarowych wykorzystywanych w systemach nadzorujących pracę urządzeń elektrycznego ogrzewania rozjazdów kolejowych

Streszczenie. *Analiza predykcyjna jest procesem wydobywania informacji z zebranych wcześniej danych, celem przewidywania przyszłych trendów i zdarzeń. Wykonane badania mają na celu zbudowanie modelu predykcyjnego wspomagającego pracę systemów urządzeń elektrycznego ogrzewania rozjazdów kolejowych (eor). W artykule opisano wykorzystanie klasyfikatorów k-NN oraz drzewa decyzyjne w celu predykcji stanu pracy urządzeń eor.*

Abstract*. Predictive analytics is the process of extracting information from previously collected data in order to anticipate future trends and events. The research carried out is aimed at building a predictive model supporting the operation of electrical heating devices for railway junctions (eor). The article describes the use of k-NN classifiers and decision trees to predict the operating state of eor devices. (**Predictive analysis of measurement data used in systems supervising the operation of electric heating devices at railway junctions**).*

Słowa kluczowe: predykcja, nagrzewanie, klasyfikator k-NN, drzewa decyzyjne.
Keywords: prediction, heating, k-NN classifier, decision tree.

Wstęp

Rozjazd kolejowy jest newralgicznym elementem drogi kolejowej, szczególnie narażonym na negatywne oddziaływanie śniegu i niskich temperatur. Najstarszym sposobem usuwania śniegu z rozjazdów, stosowanym do dnia dzisiejszego, jest ręczne oczyszczanie rozjazdów. Metoda ta może być stosowana dla linii kolejowych o niewielkich szybkościach i spotykana jest zwykle na niedużych stacjach.

Obecnie do oczyszczania przestrzeni roboczej rozjazdów kolejowych stosowane są systemy ogrzewania rozjazdów (eor), do najczęściej stosowanych rozwiązań możemy zaliczyć:

- ogrzewanie wodne obiegowe,
- ogrzewanie gazowe,
- ogrzewanie elektryczne rezystancyjne
- ogrzewanie elektryczne indukcyjne.

Urządzenia eor przeznaczone są do wytapiania śniegu i lodu z rozjazdów kolejowych w celu ich prawidłowej pracy w warunkach zimowych [1]. Do ogrzewania opornic i iglic rozjazdu przeznaczone są grzejniki płasko owalne, które są produkowane w czterech odmianach o mocy 900, 1050, 1250, 1600 W. Do ogrzewania pozostałych elementów rozjazdu stosowane są grzejniki krzyżownicowe, zamknięciowe, płyty grzewcze i inne grzejniki specjalne. W celu minimalizacji strat ciepła (szacunkowo nawet o 30%), zalecane jest stosowanie specjalnych otulin termoprzewodzących i termoizolacyjnych. Urządzenia eor są zasilane z szaf rozdzielczych poprzez skrzynie transformatorowe, które są wyposażone w transformatory eor, zabezpieczenia nadmiarowo-prądowe oraz układy tłumienia prądów udarowych.

W ostatnich latach można zauważyć znaczący wzrost łącznej mocy zainstalowanych grzejników w nowoprojektowanych urządzeniach, zarówno w powszechnie stosowanych rozjazdach typu Rz190, Rz300 i Rz500 (większa ilość zabudowanych grzałek), jak i rozjazdach przeznaczonych do jazdy z dużymi prędkościami na kierunku zwrotnym, np.: Rz1200, Rz2500 (znacząco dłuższe odcinki ruchome szyn wymagające odladzania). Monitoring i sterowanie pracą instalacji eor realizuje automat pogodowy (np. model APR-11 w przypadku urządzeń prod. AREX Sp. z o.o. - system DIMaC-EK). Automat może być zabudowany w rozdzielnicy REOR, lub na nastawni, umożliwiając tym samym dyżurnemu sterowanie pracą urządzeń. Jednocześnie pozyskane dane są na bieżąco przesyłane i archiwizowane; poprzez transmisję danych możliwy jest zdalny monitoring pracy urządzeń.

Zbudowanie modelu predykcyjnego wspomagającego pracę systemów urządzeń elektrycznego ogrzewania rozjazdów kolejowych (eor) wymaga analizy predykcyjnej danych pomiarowych. W analizie wykorzystano dwie metody klasyfikacji danych, a mianowicie drzewa decyzyjne i klasyfikator k-najbliższych sąsiadów.

Metody klasyfikacji danych

Do rozwiązania zagadnienia klasyfikacji danych wykorzystuje się obecnie algorytmy uczenia nadzorowanego i częściowo nadzorowanego dla problemów binarnych i wieloklasowych [2, 3].

Zadanie klasyfikacji sprowadza się do konstrukcji funkcji, tzw. klasyfikatora, która na podstawie zaobserwowanych cech przydziela obserwację do jednej z wcześniej zdefiniowanych grup. Estymacja klasyfikatora wykorzystuje próbę uczącą obejmującą część obserwacji.

Klasyfikacja to rodzaj nadzorowanego uczenia maszynowego, w którym algorytm „uczy się" klasyfikować nowe obserwacje na podstawie przykładowych, oznakowanych danych. Aby uzyskać większą elastyczność, można do algorytmu estymacji funkcji (klasyfikatora) przekazać np. również dane z predyktora.

Do trenowania modeli można użyć m.in. następujących klasyfikatorów [4]: drzew decyzyjnych (*decision trees*), analizy dyskryminacyjnej (*discriminant analysis*), maszyn wektorów podpierających SVM (*support vector machines*), regresji logistycznej (*logistic regression*), k najbliższych sąsiadów k-NN (*nearest neighbors*), naiwnego Bayesa (*naive Bayes*), metody Nystroema (*kernel approximation*) i sieci neuronowych (*neural network classification*).

Uczenie modelu składa się zazwyczaj z dwóch części:
• Weryfikacji modelu, polegającej na trenowaniu modelu za pomocą schematu walidacji na części danych. Algorytm trenowania powinien chronić przed nadmiernym dopasowaniem, stosując np. walidację krzyżową lub alternatywnie można wybrać wstrzymanie walidacji.

• Pełnego modelu, gdzie trenowanie modelu odbywa się na pełnych danych i jednocześnie realizowana jest weryfikacja modelu. Wynikiem jest model uczony na pełnych danych.

Kolejnym krokiem jest algorytm porównywania uzyskiwanych klasyfikatorów. Każda obserwacja z bazy danych powinna być wykorzystana do uczenia i testowania w taki sposób, aby część obserwacji wykorzystywać przy uczeniu klasyfikatorów, a pozostałe obserwacje wykorzystać do testowania. Pomocna jest tu metoda sprawdzianu krzyżowego (*cross-validation*). Kroswalidacja m-krotna czyli walidacja krzyżowa to podział danych na m części o równej liczbie obserwacji. $(m\text{-}1)$ części tworzy próbę uczącą, a m-ta część to próba testowa. Estymacja klasyfikatora następuje na próbie uczącej, a porównanie wyników na próbie testowej. Procedurę powtarza się m razy, aby każda część była próbą testową. Przed przystąpieniem do procedury walidacji krzyżowej dobrze jest przeanalizować strukturę danych i np. podzielić je na klasy. Podział próby na m części powinien uwzględniać zalecenie, aby w każdej części było po tyle samo obserwacji z każdej klasy. Ocena metody na próbie testowej może być dokonana np. poprzez estymację prawdopodobieństwa poprawnej predykcji.

Drzewa decyzyjne [5, 6]

Drzewa decyzyjne, klasyfikacyjne i regresyjne mają przewidywać odpowiedzi na dane wejściowe. Aby przewidzieć odpowiedź, należy postępować zgodnie z decyzjami w drzewie od węzła głównego (początkowego) do węzła liścia. Węzeł liścia zawiera odpowiedź. Drzewa klasyfikacyjne dają odpowiedzi nominalne, takie jak „prawda” lub „fałsz”. Drzewa regresji dają odpowiedzi liczbowe.

Klasyfikacja przy użyciu najbliższych sąsiadów k-NN [7, 8]

Kategoryzowanie punktów na podstawie ich odległości do punktów w zestawie danych treningowych może być prostym, ale skutecznym sposobem klasyfikowania nowych punktów. Do określenia odległości można użyć różnych metryk np. odległość wg Minkowskiego, Mahalanobisa czy Czebyszewa, ale najpopularniejsza jest odległość Euklidesowa.

Macierz danych oznaczono jako $\boldsymbol{X}$. Można wyznaczyć Euklidesowe odległości d_j między obiektem, będącym wektorem x i kolejno wektorami x_j następująco:

$$d_j = \left\| x - x_j \right\| \tag{1}$$

Klasyfikator k-NN należy do nieparametrycznych metod klasyfikacji. Obiekt przydzielany jest do tej klasy, do której należy większość z jego k sąsiadów ze zbioru uczącego.

Założono, że dysponujemy próbą n obiektów wylosowanych z populacji o rozkładzie $f(x)$. Prawdopodobieństwo, że k obiektów znajduje się w dostatecznie małym obszarze o objętości v przybliża funkcja:

$$f(x) = \frac{k}{n v} \tag{2}$$

Stąd estymator warunkowej gęstości w j-tej klasie wynosi:

$$f(x) = \frac{k_j}{n_j v} \tag{3}$$

gdzie: n_i – liczba obiektów z klasy j w zbiorze uczącym, k_j – liczba obiektów z j-tej klasy, które są pośród k najbliższych sąsiadów obiektu x.

Korzystając ze wzoru Bayesa [8] określono funkcje dyskryminacyjne klasyfikator k-NN:

$$g_j(x) = \frac{k_j}{k} \tag{4}$$

Na tej podstawie w [8] określono regułę decyzyjną klasyfikatora k-NN:

$$\Phi_{k-NN}(x) = i \Leftrightarrow \forall_{j=1,\ldots,k, j\neq i} g_i(x) > g_j(x) \tag{5}$$

Klasyfikator k-NN wymaga ustalenia liczby k, która zależy od liczby obserwacji w zbiorze uczącym (testowym). Estymator zgodny liczby k można zapisać wzorem

$$k = c\sqrt{n} \tag{6}$$

gdzie: c – pewna stała, $c > 0$.

Alternatywnym sposobem wyznaczenia liczby k jest jej dobór w trakcie walidacji np. metodą k-krotnej walidacji krzyżowej.

Dane pomiarowe w analizie predykcyjnej

Urządzenia eor mogą być sterowane:

• ręcznie (w trybie lokalnym lub zdalnym), sterowanie ręczne polega na bezpośrednim ręcznym załączaniu i wyłączaniu urządzeń przez obsługę posterunku ruchu, na podstawie subiektywnej oceny warunków pogodowych oraz natężenia ruchu pociągów;

• automatycznie (w trybie lokalnym lub zdalnym), sterowanie to polega na załączaniu i wyłączaniu urządzeń na podstawie warunków pogodowych analizowanych przez automat pogodowy w zależności od danych zebranych z czujników lub sterownika pogodowego (eor) [1].

Dla urządzeń eor wymagany jest poszerzony zakres przesyłania danych:

- napięcia fazowe i międzyfazowe,
- moce pobierane w poszczególnych obwodach
- temperatura szyny ogrzewanej,
- temperatura szyny nieogrzewanej,
- temperatura zamknięcia (opcja),
- wykrycie śniegu przez czujnik śniegu nawiewanego,
- wykrycie wilgoci przez czujnik wilgoci.

System monitoringu urządzeń elektroenergetyki kolejowej SMUE, pozwala na bieżący podgląd parametrów pracy urządzeń oraz umożliwia zbieranie (archiwizowanie) danych.

Dane do badań pozyskano z urządzenia pomiarowego eor na jednej ze stacji kolejowych w woj. małopolskim. Analizę prowadzono na zebranych danych pomiarowych (dane źródłowe) z okresu 30 dni okresu zimowego – styczeń 2013 r. Pozyskano w ten sposób blisko 12 tysięcy (rekordów) odczytów.

Dane te, przedstawione na rysunku 1, mają charakter „mieszany” w postaci danych źródłowych i kodów, obejmujących:

• data i godzina (kolumny A i B), przeliczono te odczyty na narastające odstępy czasowe (kolumna C), a następnie wyliczono różnice pomiędzy poszczególnymi odczytami sprowadzając je do jednolitych jednostek (sekund – kolumna H),

• dopisano z danych pomiarowych sterownika rozdzielnicy parametr załączenia grzałek; parametr binarny 0 / 1 (kolumna D),

• dane dotyczące temperatury zewnętrznej, temperatury szyny zimnej i temperatury szyny ogrzanej podane są w st. Celsjusza (kolumny E, F, G),

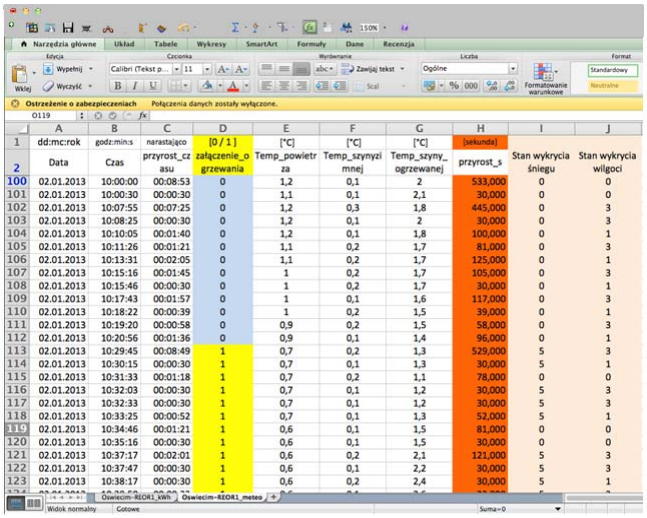
• stan wykrycia opadów śniegu oraz wykrycia wilgoci (kolumny I oraz J) podane są w formie liczbowej w zakresie od 0 do 7; gdzie poszczególne cyfry oznaczają:

„0” oznacza brak wykrytego śniegu,

„1” to śnieg wypadkowy, zawiera w sobie wszystkie pomiary własne i nadesłane przez sterownik nadrzędny rozdzielnicy,
„2” to śnieg wykryty przez przetwornik pogodowy,
„4” to śnieg wykryty na podstawie wilgoci i progu temperatury,
cyfry „3”, „5”, „6” i „7” są odpowiednio sumą arytmetyczną ww. cyfr „0”, „1”, „2” i „4” np. wartość „3” to śnieg wykryty z powodów „1” i „2”.

Tak duża złożoność i różnorodność danych wymagała ich dalszej obróbki oraz sprowadzenia „do wspólnego mianownika”, co umożliwiało ich dalszą obróbkę:

- nieregularne odczyty, za pomocą stworzonego algorytmu podzielono na odczyty o jednakowych jednosekundowych odstępach czasowych,
- dane o wykryciu śniegu i wilgoci zamieniono na postać binarną.



	A	B	C	D	E	F	G	H	I	J
1	dd:mc:rok	godz:min:s	narastająco	[0 / 1]	[°C]	[°C]	[°C]	[sekunda]		
2	Data	Czas	przyrost_czasu	załączenie_ogrzewania	Temp_powietrza	Temp_szynyzimnej	Temp_szyny_ogrzewanej	przyrost_s	Stan wykrycia śniegu	Stan wykrycia wilgoci
100	02.01.2013	10:00:00	00:08:53	0	1,2	0,1	2	533,000	0	0
101	02.01.2013	10:00:30	00:00:30	0	1,1	0,1	2,1	30,000	0	0
102	02.01.2013	10:07:55	00:07:25	0	1,2	0,3	1,8	445,000	0	3
103	02.01.2013	10:08:25	00:00:30	0	1,2	0,1	2	30,000	0	3
104	02.01.2013	10:10:05	00:01:40	0	1,2	0,1	1,8	100,000	0	1
105	02.01.2013	10:11:26	00:01:21	0	1,1	0,2	1,7	81,000	0	3
106	02.01.2013	10:13:31	00:02:05	0	1,1	0,2	1,7	125,000	0	1
107	02.01.2013	10:15:16	00:01:45	0	1	0,2	1,7	105,000	0	3
108	02.01.2013	10:15:46	00:00:30	0	1	0,2	1,7	30,000	0	1
109	02.01.2013	10:17:43	00:01:57	0	1	0,1	1,6	117,000	0	3
110	02.01.2013	10:18:22	00:00:39	0	1	0,2	1,5	39,000	0	1
111	02.01.2013	10:19:20	00:00:58	0	0,9	0,2	1,5	58,000	0	3
112	02.01.2013	10:20:56	00:01:36	0	0,9	0,1	1,4	96,000	0	1
113	02.01.2013	10:29:45	00:08:49	1	0,7	0,2	1,3	529,000	5	3
114	02.01.2013	10:30:15	00:00:30	1	0,7	0,1	1,3	30,000	5	1
115	02.01.2013	10:31:33	00:01:18	1	0,7	0,2	1,1	78,000	0	0
116	02.01.2013	10:32:03	00:00:30	1	0,7	0,1	1,2	30,000	5	3
117	02.01.2013	10:32:33	00:00:30	1	0,7	0,1	1,2	30,000	5	3
118	02.01.2013	10:33:25	00:00:52	1	0,7	0,1	1,3	52,000	5	1
119	02.01.2013	10:34:46	00:01:21	1	0,6	0,1	1,5	81,000	0	0
120	02.01.2013	10:35:16	00:00:30	1	0,6	0,1	1,5	30,000	0	0
121	02.01.2013	10:37:17	00:02:01	1	0,6	0,2	2,1	121,000	5	3
122	02.01.2013	10:37:47	00:00:30	1	0,6	0,1	2,2	30,000	5	3
123	02.01.2013	10:38:17	00:00:30	1	0,6	0,2	2,4	30,000	5	1

Rys.1. Przykładowe dane pomiarowe wykorzystane w analizie (Źródło: opracowanie własne)

W trakcie prowadzonej analizy danych pomiarowych zastosowano dwuetapowy algorytm:

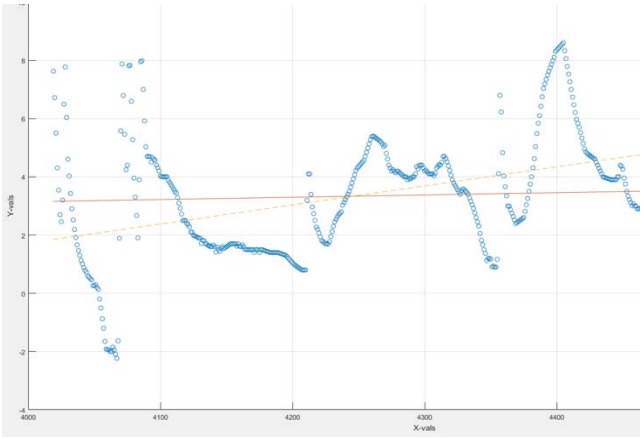
- wykorzystano *metody regresji liniowej*, celem przewidzenia ogólnego trendu analizowanych danych,
- wykorzystano *klasyfikatory k-NN oraz drzewa decyzyjne*, celem przewidzenia odpowiedzi systemu.

Zastosowanie metody regresji liniowej pozwoliło na oszacowanie ogólnego trendu danych. Środowisko Matlab umożliwia nam stosowanie odmiennych metod regresji oraz testowania jakości oszacowanych danych. Przykład zastosowania metody regresji liniowej przedstawiono na rysunku 2.

Przed przeprowadzeniem klasyfikacji, istotnym zagadnieniem jest kwestia właściwego przetworzenia danych (redukcja liczności zbioru, transformaty), co może znacząco przyśpieszyć obliczenia i zwiększyć celność (*accurancy*) algorytmu.

W celu uniknięcia nadmiernego przystosowania (przeuczenia) klasyfikatora do danych (tzw. *overfitting*), zastosowano metodę sprawdzianu krzyżowego (*cross-validation*). Polegało to na podziale próby statystycznej na k podzbiorów, a następnie przeprowadzeniu wszelkich analiz na niektórych z nich – każdy z nich bierze się jako testowy, a wszystkie pozostałe jako zbiór uczący. Otrzymane rezultaty następnie uśrednia się i w ten sposób otrzymuje się klasyfikator lepszy jakościowo.

Teoria sprawdzianu krzyżowego pozwala bronić się przez tzw. błędem trzeciego rodzaju i właściwie ocenić trafność diagnostyczną modelu predykcyjnego. Bez zastosowania metody sprawdzianu krzyżowego nie można mieć pewności czy model będzie dobrze działał dla danych, które nie były wykorzystywane do jego konstruowania.



Rys.2. Dane temperaturowe z urządzeń pomiarowych wraz z linią regresji (linia ciągła kolor pomarańczowy) oraz stałą regresji (Y-intercept, linia przerywana), (Źródło: opracowanie własne)

Dzięki zastosowaniu klasyfikacji nadzorowanej, jesteśmy w stanie określić, do jakiej klasy należy określona próbka danych, co pozwala nam na interpretowanie wyników regresji. Klasyfikatory pozwalają nam z pewna celnością odróżnić, w których momentach urządzenia grzejne powinny być załączone, a w których nie. Wykorzystanie *toolbox*'ów wbudowanych w Matlaba pozwala nam na przeprowadzenie wielu modeli klasyfikacyjnych i wybór właściwego, który zapewni nam najlepszą jakość, na którą składają się kryteria celności (*accurancy*), poziomu nadmiernego dopasowania (overfitting) oraz czasu szkolenia.

Zastosowania klasyfikatorów k-NN i drzew decyzyjnych w analizie danych pomiarowych eor

Najistotniejszym zadaniem klasyfikacji jest budowa określonego modelu, który służy do predykcji przydziału do klasy prognozowanej zmiennej. Wybranymi przykładami metod nadzorowanych (uczenie maszynowe) są:

- algorytmy k-NN (k-najbliższych sąsiadów, $k=5$) – metoda hierarchiczna [8],
- drzewa decyzyjne – metoda niehierarchiczna [6].

W algorytmie k-NN dane pomiarowe dzieli się na zbiór testowy i zbiór treningowy. Klasyfikator opierając się na zbiorze treningowym uczy się właściwości danych i przypisuje każdemu wektorowi klasę, czyli wartość decyzyjną, będącą wielkością wyjściową modelu. Dokładność klasyfikatora wyznaczana jest poprzez porównanie wartości decyzyjnych ze zbioru testowego z klasami przewidzianymi przy użyciu tego modelu [2-5].

W celu uniknięcia nadmiernego przystosowania (przeuczenia) klasyfikatora do danych (tzw. *overfitting*), zastosowano tu metodę sprawdzianu krzyżowego (*cross-validation*), polegającego na podziale próby statystycznej na $k=5$ podzbiorów.

Dzięki zastosowaniu klasyfikacji nadzorowanej określono, do jakiej klasy należy określona próbka danych, co pozwala na interpretowanie wyników analizy. Klasyfikatory pozwalają nam z pewną celnością odróżnić, w których momentach urządzenia grzejne eor powinny być załączone, a w których nie.

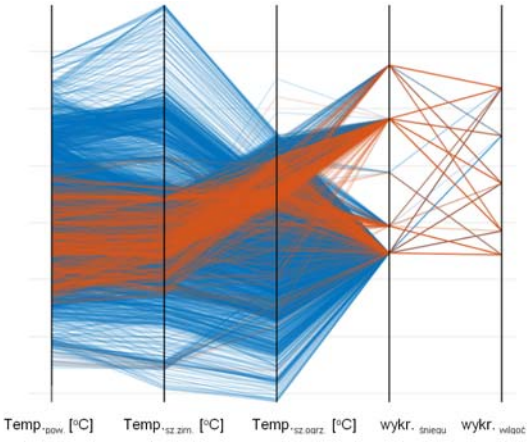
Drzewo decyzyjne (klasyfikacyjne) reprezentuje proces podziału zbioru obiektów na jednorodne klasy. Drzewo jest grafem, w którym każdy wewnętrzny węzeł odpowiada pewnej hipotetycznej decyzji, a elementy końcowe (liście) reprezentują cele (klasy, do których należą obiekty) [6].

Drzewa decyzyjne są istotnym narzędziem w uczeniu maszynowym i eksploracji danych, są one powszechnie wykorzystywane do rozwiązywania zadań klasyfikacji.

W analizie zastosowano metodę CART (Classification and Regression Trees) wykorzystywaną przez funkcje w pakiecie Matlab, a konkretnie utworzono drzewo klasyfikacyjne służące do przyporządkowania klas charakteryzujących zmienną docelową. Stosowano regułę podziału na dwie części. Natomiast w każdym węźle drzewa wyszukiwano podział, który daje najlepszą trafność predykcji. Procedury Matlaba umożliwiają automatyczną optymalizację hiperparametrów (np. minimalną liczbę próbek w liściu).

Rezultaty analizy

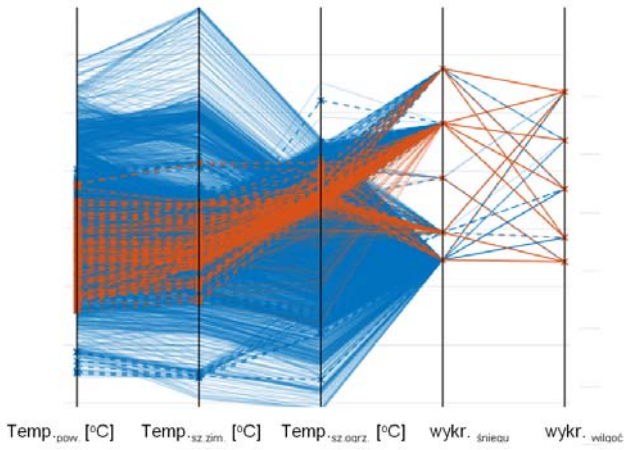
Na rysunku 3 przedstawiono dane pomiarowe na wykresie współrzędnych równoległych.



Rys.3. Reprezentacja danych na wykresie współrzędnych równoległych (Źródło: opracowanie własne)

Legenda:
- kolor pomarańczowy – klasa „1” (grzanie włączone),
- kolor niebieski – klasa „0” (grzanie wyłączone),
- czarne pionowe linie (słupki)- (kolejno od lewej): wartość temperatury powietrza, wartość temperatury szyny zimnej, wartość temperatury szyny ogrzewanej, wykrycie śniegu, wykrycie wilgoci.

Przykłady klasyfikacji z zastosowaniem klasyfikatorów k-NN przedstawiono na rysunku 4, natomiast na rysunku 5 przedstawiono klasyfikację przy użyciu drzewa decyzyjnego.

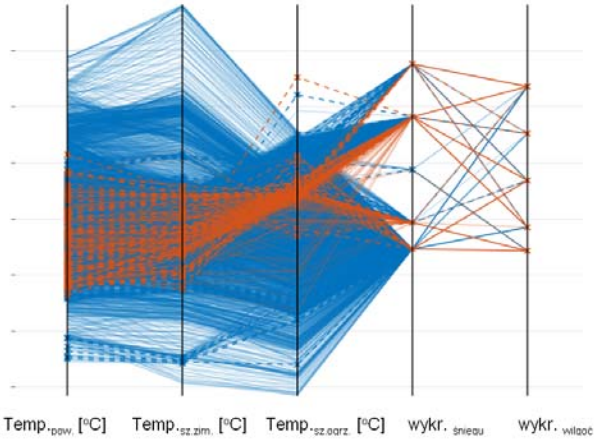


Rys.4. Klasyfikacja danych przy użyciu klasyfikatora k-NN (k-najbliższych sąsiadów) (Źródło: opracowanie własne)

Uzyskana dokładność klasyfikacji metodą k-NN wynosi 87,4 %, a przy użyciu drzewa decyzyjnego 88,4 %.

Wnioski i uwagi końcowe

Algorytm k-NN najbliższych sąsiadów jest użyteczny szczególnie wtedy, gdy zależność między zmiennymi objaśniającymi a objaśnianymi jest złożona lub nietypowa (np. niemonotoniczna), czyli trudna do modelowania w klasyczny sposób. W badaniach ustawiono k-krotny ($k=5$) sprawdzian krzyżowy, co pozwoliło na uzyskanie klasyfikatorów nie idealnie dopasowanych do danych, lecz do ogólnego trendu.



Rys.5. Klasyfikacja danych przy użyciu klasyfikatora drzewiastego (drzewa decyzyjne) (Źródło: opracowanie własne)

W analizie predykcyjnej uzyskano dużą zbieżność rezultatów klasyfikacji metodą k-NN i klasyfikacji drzewa decyzyjnego do danych pomiarowych (uzyskana dokładność klasyfikacji odpowiednio 87,4 % i 88,4 %). Potwierdza to zasadność zastosowania obu metod do klasyfikacji danych i daje podstawę do zastosowania obu metod w badaniach do wykonania predykcji oraz wykorzystania jej w modelach decyzyjnych.

***Autorzy**: mgr inż. Artur Wachtarczyk, doktorant, Politechnika Częstochowska; dr hab. inż. Janusz Sowiński, prof. uczelni, Politechnika Częstochowska, Katedra Elektroenergetyki, Al. Armii Krajowej 17,42-200 Częstochowa, E-mail: janusz.sowinski@pcz.pl;*

LITERATURA

[1] Wytyczne do projektowania urządzeń elektrycznego ogrzewania rozjazdów PKP Polskie Linie Kolejowe S.A. – Instrukcja Iet-5, Warszawa, (2015)
[2] Suthaharan S., *Machine Learning Models and Algorithms for Big Data Classification*, (2015), Springer New York, NY
[3] Everitt B. S., Landau S., Leese M., Stahl D., *Miscellaneous Clustering Methods*, in *Cluster Analysis*, 5th Edition (2011), John Wiley & Sons, Ltd., Chichester, UK
[4] Aggarwal C.C. (Ed.), *Data Classfication. Algorithms and Application*, (2020), Chapman & Hall/CRC
[5] Nguyen H.S., Systemy decyzyjne, Uniwersytet Warszawski, 2011, https://mst.mimuw.edu.pl/lecture.php?lecture=syd
[6] Fürnkranz J., Decision Tree, in: Sammut C., Webb G.I. (Eds.) *Encyclopedia of Machine Learning*, (2011), Springer, Boston, MA, pp 263–267
[7] Zhang S., Li X., Zong M., Zhu X., Wang R., Efficient kNN Classification With Different Numbers of Nearest Neighbors, in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5 (2018), pp. 1774-1785
[8] *K-nearest neighbors algorithm*, Wikipedia, https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm (access 25 April 2023)