

Mgr Mateusz Wyrembek  
 Uniwersytet Ekonomiczny w Poznaniu  
 ORCID: 0000-0002-7946-948X  
 e-mail: Mateusz.Wyrembek@phd.ue.poznan.pl

# Zastosowanie metod uczenia maszynowego do przewidywania ryzyka uszkodzeń towarów

*The use of machine learning methods to predict the risk of damage to goods*

## Streszczenie

Zapotrzebowanie transportowe z roku na rok zwiększa się. Dzieje się tak z powodu wzrostu produkcji i konsumpcji. Czym dłuższy jest łańcuch dostaw, tym większe są możliwości wystąpienia w nim zakłóceń, gdyż wszystkie operacje wiążą się z ryzykiem. Jednakże o ryzyku w kontekście łańcucha dostaw i zarządzania nim mówi się od niedawna. Niestety żadna branża nie jest odporna na przewidywalne i nieprzewidywalne zakłócenia, które wywołują straty (np. zaginięcie towaru). Z punktu widzenia przewoźników istotna byłaby możliwość przewidzenia wystąpienia np. uszkodzenia towaru. Artykuł koncentruje się na wykorzystaniu metod uczenia maszynowego do przewidzenia ryzyka uszkodzeń towarów (takich jak sprzęt RTV, AGD czy telefony/komputery) w transporcie drogowym. W ramach badań wykorzystano pięć inteligentnych metod: regresję logistyczną, maszynę wektorów nośnych (SVM); drzewo decyzyjne; naiwny klasyfikator bayesowski; AdaBoost. Natomiast celem jest przedstawienie koncepcji oraz wymienionych metod uczenia maszynowego, miar oceniających wydajność modeli oraz wyników związanych z przeprowadzonym badaniem. Postawiony cel zdeterminował wybór metod badawczych – wykorzystano analizę literatury oraz programowanie. W ostatniej części artykułu przedstawiono wyniki otrzymane z analizy pięciu modeli. W ramach badań ustalono, że najlepszą zdolność predykcyjną ma AdaBoost.

## Słowa kluczowe:

uczenie maszynowe; zarządzanie ryzykiem

## Abstract

The transport demand is increasing year by year. This is because of the increase in production and consumption. The longer the supply chain is, the more likely it is to be disrupted, as all operations involve risks. However, risk in the context of the supply chain and its management has recently been discussed. Unfortunately, no industry is immune to predictable and unpredictable disruptions that affect losses (e.g. loss of goods). From the point of view of carriers, it would be important to be able to predict the occurrence of, for example, damage to the goods. The article focuses on the use of machine learning methods to predict the risk of damage to goods (such as electronics, household appliances or telephones/computers) in road transport. The research used five intelligent methods such as: logistic regression; support vector machine (SVM); decision tree; naive Bayesian classifier; AdaBoost. The aim of the paper is to present the concept and the above-mentioned methods of machine learning, measures assessing the performance of models and the results related to the conducted research. The set goal determined the choice of the research methods – literature analysis and programming were used. The last part of the article presents the results obtained from the analysis of five models. The research established that AdaBoost has the best predictive ability.

## Keywords:

machine learning, risk management

JEL: R41, C45, C53, D81

## Wstęp

Zapotrzebowanie transportowe z roku na rok zwiększa się. Dzieje się tak z powodu wzrostu produkcji i konsumpcji. W wyniku globalizacji przedsiębiorstwa lokalizują swoją działalność produk-

cyjną z dala od rynku zbytu, co ma wpływ na wydłużenie się łańcuchów logistycznych (Anholcer, 2019, s. 11). Czym dłuższy jest łańcuch dostaw, tym większe są możliwości wystąpienia w nim zakłóceń, gdyż wszystkie operacje gospodarcze wiążą się z ryzykiem.

Jednakże o ryzyku w kontekście łańcucha dostaw i zarządzania nim mówi się od niedawna. Zauważają to G. Baryannis i in. (2018), którzy stwierdzili, że idea zarządzania ryzykiem w łańcuchu dostaw (SCRM) zaczęła przyciągać uwagę ze względu na niepewność gospodarczą, która rozpoczęła się wraz z kryzysem finansowym w 2008 r. D. Waters (2007) definiuje SCRM jako koncepcję, która obejmuje identyfikację, analizę i kontrolę ryzyka w łańcuchu dostaw w celu zwiększenia jego odporności.

Niestety żadna branża nie jest odporna na przewidywalne i nieprzewidywalne zakłócenia, które wywołują straty (np. zaginięcie towaru). Z punktu widzenia przewoźników istotna byłaby możliwość przewidzenia wystąpienia np. uszkodzenia towaru, gdyż zgodnie z krajowymi i międzynarodowymi przepisami odpowiadają oni za szkody wynikłe od momentu przyjęcia aż do czasu wydania przesyłki. W związku z tym badacze, w ramach usprawnień decyzji i działań w kontekście zarządzania ryzykiem, zaproponowali wykorzystanie m.in. metod uczenia maszynowego (Baryannis i in., 2019). Pomimo sporej liczby prac z tego obszaru nie udało się znaleźć opracowań poświęconych wykorzystaniu inteligentnych algorytmów do predykcji uszkodzeń towarów w transporcie. Istnieje zatem pewna luka badawcza, która zmotywowała autora do podjęcia prób jej uzupełnienia.

Artykuł koncentruje się na wykorzystaniu metod uczenia maszynowego do przewidzenia ryzyka uszkodzeń towarów (takich jak sprzęt RTV, AGD czy telefony/komputery) w transporcie drogowym. W ramach badań wykorzystano pięć inteligentnych metod, a mianowicie:

- regresję logistyczną;
- maszynę wektorów nośnych (SVM);
- drzewo decyzyjne;
- naiwny klasyfikator bayesowski;
- AdaBoost.

Natomiast celem opracowania jest przedstawienie koncepcji oraz wskazanych metod uczenia maszynowego, miar oceniających wydajność modeli oraz wyników związanych z przeprowadzonym badaniem.

## Uczenie maszynowe

W literaturze można się spotkać z wieloma różnymi definicjami uczenia maszynowego. Przykładowo H. Wenzel i in. (2019) definiują uczenie maszynowe jako: „podobszar sztucznej inteligencji, który reprezentuje inny sposób programowania. Przykładowe dane zastępują sztywne reguły obliczeniowe programu. Z podanych przykładowych danych metody lub algorytmy uczenia wydobywają prawidłowości statystyczne i przedstawiają je w postaci mo-

deli. Modele mogą reagować na nowe, nieznane dane i klasyfikować je do kategorii lub dokonywać predykcji”. Natomiast według T. Mittchela (1997) uczenie maszynowe polega na tym, że: „Program komputerowy (maszyna) uczy się na podstawie doświadczenia ( $E$ ) w odniesieniu do pewnej klasy zadań ( $T$ ) i miary efektywności ( $P$ ), jeśli jego efektywność wykonywania zadania  $T$  (mierzona za pomocą  $P$ ) poprawia się wraz z doświadczeniem  $E$ ”. Stwierdzenie to zostało przedstawione na rysunku 1.

Do programu implementowane są dane wyjściowe (zmienna objaśniająca) wraz ze spodziewanymi wynikami (zmienna objaśniana) – patrz pkt E na rysunku 1. Rolą systemu jest poznanie reguł generowanych przez zaobserwowane rezultaty poszczególnych danych wyjściowych. Oprogramowanie trenuje model, który przedstawia to, czego nauczył się do tej pory z danych treningowych. Po fazie uczenia system jest przygotowany do wnioskowania. Implementując nowe dane wyjściowe (dane walidacyjne), przeszkolony model może przedstawić oczekiwany wynik (pkt T). Zestaw danych testowych, przewidzianych oraz oczekiwanych wyników określa efektywność przewidywania – pkt P (Laska, 2019, s. 23).

Warto również zwrócić uwagę na trzy podstawowe rodzaje uczenia maszynowego (przedstawione na rysunku 2), które definiuje się następująco:

- **uczenie nadzorowane** – algorytm tworzy funkcję, która na podstawie zmiennych wejściowych  $x$  oraz zmiennej wyjściowej  $y$  przewiduje przyszły wynik (Cunningham i in., 2008);
- **uczenie nienadzorowane** – odnosi się do algorytmów identyfikujących wzorce w zestawach danych, które nie są ani sklasyfikowane, ani oznakowane (Dridi, 2021, s. 1);
- **uczenie przez wzmacnianie** – zachodzi, kiedy system działa w środowisku nieznanym – nie są określone dane wejściowe ani wyjściowe; jedyną informacją, jaką otrzymuje maszyna, jest sygnał wzmocnienia, który może być pozytywny bądź negatywny (Laska, 2019, s. 38).

## Metody uczenia maszynowego

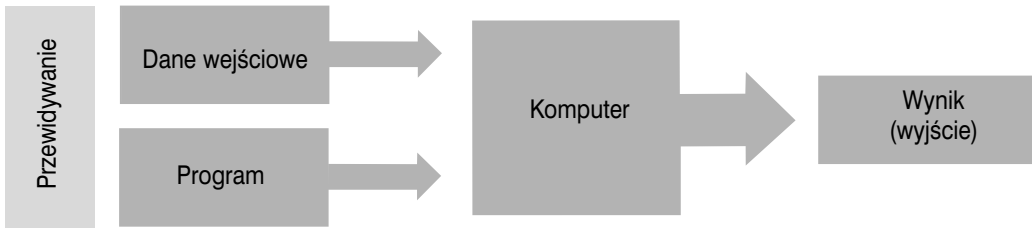
Jak już zostało wcześniej wspomniane, w badaniu wykorzystano pięć metod uczenia maszynowego: regresję logistyczną, maszynę wektorów nośnych, drzewo decyzyjne, naiwny klasyfikator bayesowski oraz AdaBoost, należących do metod uczenia nadzorowanego. Mając to na uwadze, tylko te metody zostały zdefiniowane.

**Regresja logistyczna** służy do oszacowania związku jednej lub większej liczby zmiennych niezależnych (predykcyjnych) z binarną zmienną zależną. Innymi słowy metoda ta jest wykorzystywana do oszacowania prawdopodobieństwa lub ryzyka kon-

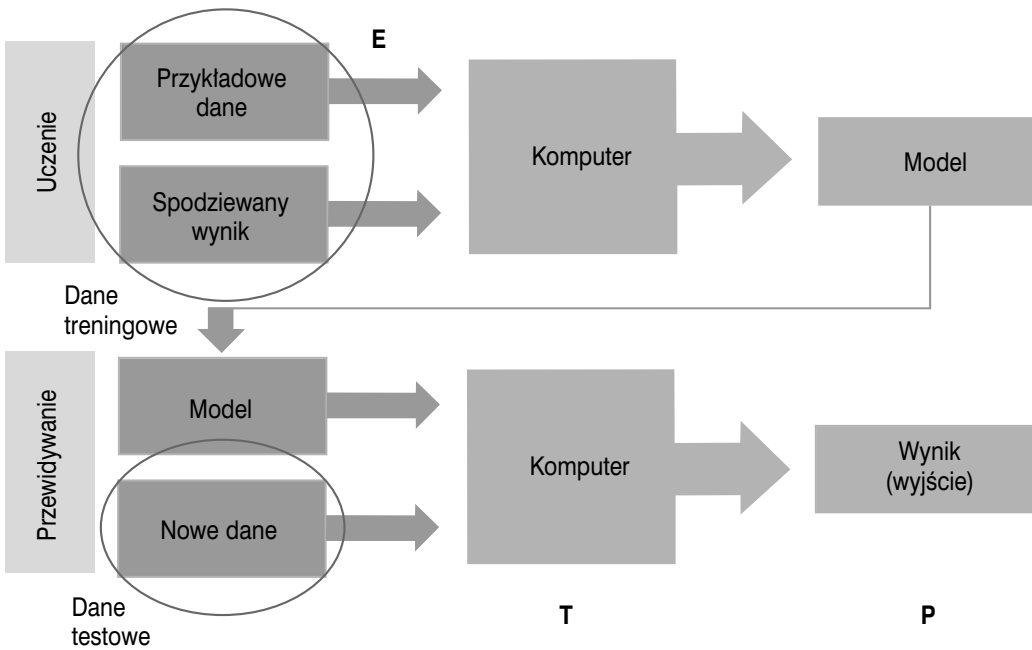
Rysunek 1

Graficzne przedstawienie kontrastu pomiędzy tradycyjnym programowaniem a uczeniem maszynowym (tu: nadzorowanym)

**Tradycyjne oprogramowanie:**



**Uczenie maszynowe**



Źródło: Laska, 2019, s. 23

Rysunek 2

Rodzaje uczenia maszynowego



Źródło: opracowanie własne.

## Rysunek 3

## Macierz pomyłek

	Wynik pozytywny	Wynik negatywny
Naprawdę pozytywny	Prawidłowy wynik pozytywny ( <i>true positive</i> , TP)	Błędny wynik pozytywny ( <i>false positive</i> , FP)
Naprawdę negatywny	Prawidłowy wynik negatywny ( <i>true negative</i> , TN)	Błędny wynik negatywny ( <i>false negative</i> , FN)

Źródło: Joachimiak, 2016, s. 31.

kretnego wyniku, biorąc pod uwagę wartości zmiennych niezależnych (Schober & Vetter, 2021).

**Maszyna wektorów nośnych** jest metodą uczenia maszynowego używaną do klasyfikacji oraz regresji. W przypadku regresji celem jest znalezienie funkcji  $f$ , dla której błąd prognozowania w odniesieniu do celu  $y$  jest co najwyżej równy wstępnie zdefiniowanemu progowi tolerancji  $\varepsilon$  dla elementów zbioru uczącego (Drucker i in., 1999).

**Drzewo decyzyjne** jest techniką opartą na schemacie drzewa, w którym każda ścieżka rozpoczynająca się od korzenia jest opisana sekwencją oddzielających dane, aż do osiągnięcia wyniku logicznego w węźle liścia. Model ma na celu łączenie serii podstawowych testów, gdzie cecha numeryczna jest porównywana z wartością progową w każdym teście. Warto zaznaczyć, że każdy węzeł reprezentuje cechy w kategorii, która ma być sklasyfikowana, a każdy podzbiór definiuje wartość, którą może przyjąć węzeł (Su & Zhang, 2006).

**Naiwny klasyfikator bayesowski** to prosty klasyfikator, który oblicza zbiór prawdopodobieństwa na podstawie częstotliwości i kombinacji wartości w danym zbiorze danych. Algorytm wykorzystuje twierdzenie Bayesa i zakłada, że wszystkie zmienne są niezależne, biorąc pod uwagę wartość zmiennej klasy. Klasyfikator nazywany jest naiwnym ze względu na rzadkość sprawdzania się przyjętego założenia w rzeczywistości. Jednakże należy wspomnieć, że algorytm potrafi szybko się uczyć (Sartias & Yasar, 2019, s. 90).

**AdaBoost** to algorytm opierający się na boostingu. Metoda ta polega na połączeniu wielu różnych klasyfikatorów, aby otrzymać jeden lepszy. AdaBoost buduje klasyfikator, na przykład drzewo klasyfikacyjne, które wytwarza etykiety klas. Jeśli treningowy punkt danych jest błędnie sklasyfikowany, waga tego punktu danych zostaje zwiększona (Zhu i in., 2009, s. 350).

## Ewaluacja modeli

Każdy model należy poddać ocenie predykcji, czyli sprawdzić pomiar wydajności. Innymi słowy powinno się zweryfikować, jak skutecznie zbudowany system przewiduje nowe dane, czyli takie, które nie biorą udziału w procesie uczenia. Narzędziem do tego służącym jest macierz pomyłek (przedstawiona na rysunku 3), czyli tablica  $N \times N$ , gdzie  $N$  to liczba różnych klas (Visa i in., 2011). Należy przyjąć, że kolumny reprezentują klasy rzeczywiste, natomiast wiersze klasy predykcyjne.

W przypadku klasyfikacji binarnej ryzyka uszkodzeń towaru można wyróżnić dwa rodzaje błędów – pozytywne oraz negatywne, gdzie:

- **prawidłowy wynik pozytywny (TP)** – model prawidłowo przewidział wystąpienie ryzyka uszkodzenia towaru;
- **prawidłowy wynik negatywny (TN)** – model prawidłowo przewidział brak ryzyka uszkodzenia towaru;
- **błędny wynik pozytywny (FP)** – model błędnie przewidział wystąpienie ryzyka uszkodzenia towaru;
- **błędny wynik negatywny (FN)** – model błędnie przewidział brak ryzyka uszkodzenia towaru.

Na podstawie macierzy pomyłek można wyliczyć miary wydajności modelu, m.in. takie jak (Iwendi i in., 2022; Joachimiak, 2016):

- **dokładność (*accuracy*)** – to procent odpowiadający wszystkim prawidłowym odpowiedziom klasyfikatora; jest określana wzorem:

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN};$$

- **precyzja (*precision*)** – mierzy skuteczność, jest to zatem stosunek prawidłowych odpowiedzi do wszystkich pozytywnych wyników; jest określana wzorem:

$$precision = \frac{TP}{TP+FP};$$

- **pokrycie (recall)** – jest to stosunek prawidłowego wyniku pozytywnego do sumy wyników pozytywnych oraz błędnych wyników negatywnych; jest określane wzorem:

$$recall = \frac{TP}{TP+FN};$$

- **miara F1** – miara ta jest średnią precyzji oraz pokrycia; jest obliczana jest według wzoru:

$$F1 = \frac{2 \times precision \times recall}{precision + recall}.$$

Do ewaluacji modeli są również wykorzystywane miernik AUC (*Area Under Curve*) i krzywa ROC. Pierwsza z miar określa trafność modelu przy przyjęciu wartości z przedziału [0; 1]. Im bliżej wartości 1, tym lepszą zdolność predykcyjną ma model. Krzywa ROC natomiast obrazuje wydajność (Wyrembek, 2022, s. 35).

## Ekspertyzy na zaimplementowanych modelach uczenia maszynowego

Do eksperymentu użyto zbioru danych pochodzących z jednego z ogniw łańcucha dostaw przedsiębiorstwa zajmującego się dystrybucją (organizacja trudni się dystrybucją sprzętu AGD, RTV, telefonów i komputerów). Informacje te pochodzą z systemu TMS. Badany zbiór zawiera 58 350 obserwacji w 16 wierszach. Ekspertyzy przeprowadzono z wykorzystaniem sprzętu komputerowego przedstawionego w tabeli 1. Głównym językiem programowania był język R. Zgodnie z przedstawioną już definicją T. Mitchela przed przystąpieniem do eksperymentu wyodrębniono trzy składowe zadania uczenia maszynowego, które można zdefiniować następująco:

- zadanie *T* – rozpoznanie i sklasyfikowanie ryzyka uszkodzenia towaru;
- miara wydajności *P* – procent poprawnie sklasyfikowanych uszkodzeń;
- doświadczenie szkoleniowe *E* – baza danych paczek z 2021 r. oraz poprawnie określony status szkody.

Na podstawie mapy braków ustalono, że brakuje 3% wartości w kategorii numer WZ (kategoria nie będzie wykorzystywana w badaniu).

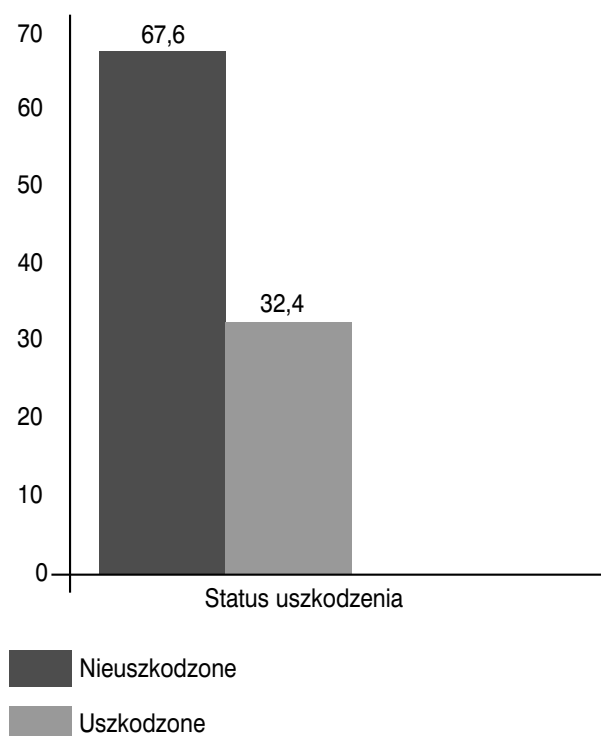
Rysunek 4 przedstawia procent uszkodzonych oraz nieuszkodzonych paczek w 2021 r. Można zauważyć, że aż 32,4% paczek uległo uszkodzeniu. Natomiast na

Tabela 1  
Hardware używany do przeprowadzenia eksperymentów

Sprzęt	Specyfikacja
System operacyjny	macOS Monterey 12.5
Procesor (CPU)	8-rdzeniowy CPU
Procesor graficzny (GPU)	7-rdzeniowy GPU
Pamięć komputera	8 GB

Źródło: opracowanie własne.

Rysunek 4  
Paczki uszkodzone oraz nieuszkodzone w 2021 r. (w %)



Źródło: opracowanie własne.

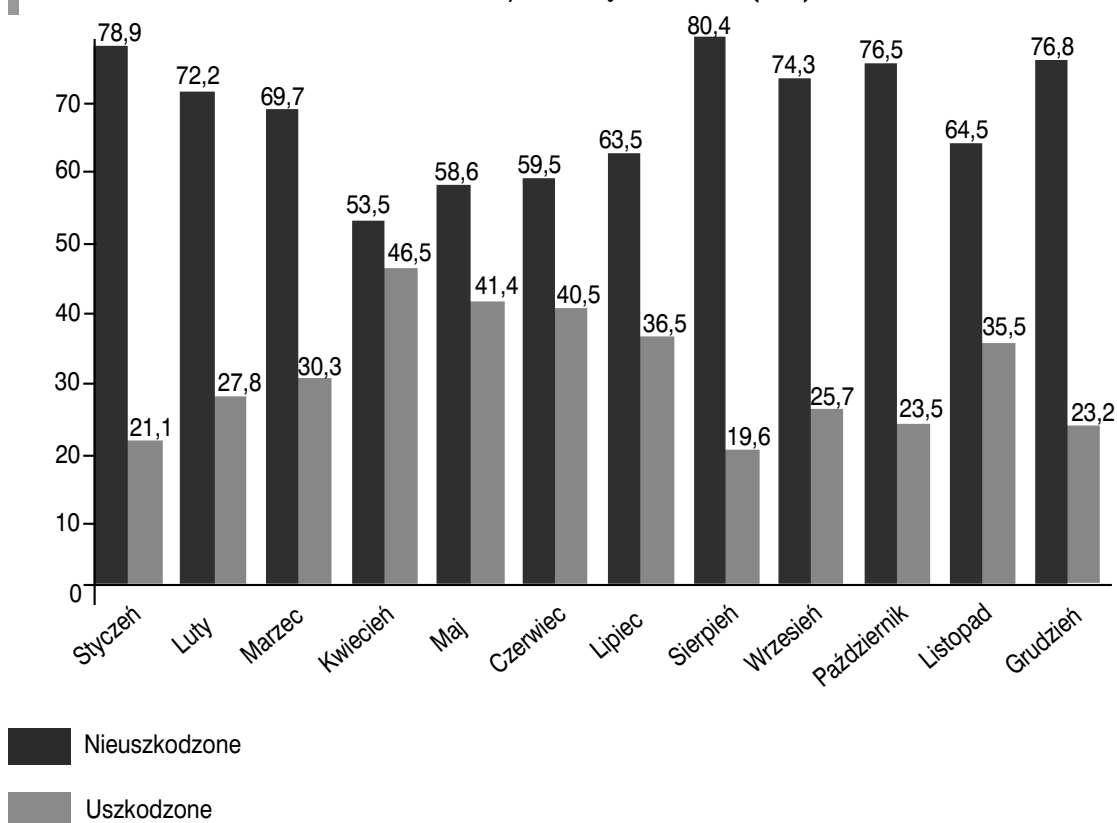
rysunku 5 zobrazowano rozkład uszkodzeń w ujęciu miesięcznym. Najmniej towarów uszkodziło się w sierpniu, natomiast najwięcej w kwietniu.

Do eksperymentu wybrano kategorie zmiennych określone w tabeli 2. Zestaw danych został podzielony losowo na dwie części – treningową (70% danych) oraz walidacyjną (30%).

Na wybranych danych zastosowano nadzorowane metody uczenia maszynowego. W tabeli 3 przed-

Rysunek 5

Paczki uszkodzone oraz nieuszkodzone w danym miesiącu w 2021 r. (w %)



Źródło: opracowanie własne.

Tabela 2

Kategorie wybrane do eksperymentu

Kategoria	Specyfikacja
Miesiąc	Dany miesiąc, w którym odbywał się transport towaru
Kwartał	Dany kwartał, w którym odbywał się transport towaru
Terminal	Terminal odbierający towar od nadawcy, w którym odbywa się przeładunek do terminalu wysyłającego
Towar	Rodzaj przewożonego towaru
Status szkody	Status uszkodzenia określony binarnie: 0 – towar nieuszkodzony; 1 – towar uszkodzony
Miasto	Miasto, do którego miał być dostarczony towar
Kraj	Kraj, do którego towar miał zostać dostarczony
Terminal wysyłający	Terminal, z którego towar jest dostarczany do odbiorcy
Usługa	Typ usługi: express – dostawa w ciągu 48 godz.; standard – dostawa w ciągu 72 godz.
Partner	Przewoźnik, który przewozi towar z terminalu odbierającego do wysyłającego

Źródło: opracowanie własne.

stawiono wyniki badań (najlepsze wyniki zostały pogrubione). Na ich podstawie należy stwierdzić, że każdy z modeli charakteryzuje się dość dobrą predykcją. Jednak zdecydowanie najlepszą moc

posiada AdaBoost. Również krzywa ROC potwierdza wysoką wydajność modelu (krzywe ROC dla wszystkich modeli zostały przedstawione na rysunku 6a–e).

Tabela 3

Wyniki przeprowadzonych badań

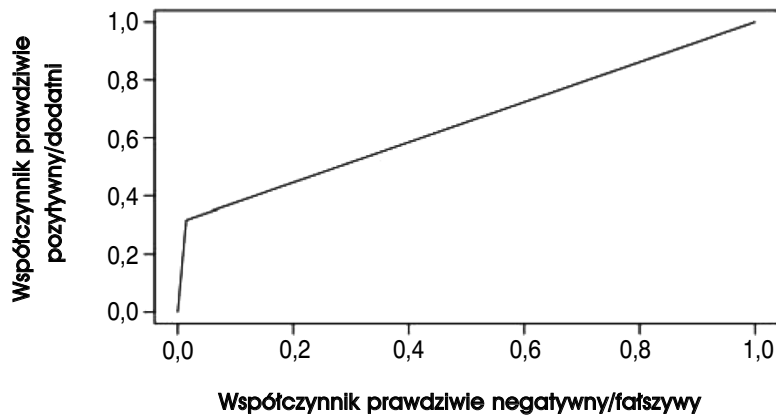
Miara/metoda	Maszyna wektorów nośnych	Drzewo decyzyjne	Naiwny klasyfikator bayesowski	AdaBoost	Regresja logistyczna
Accuracy	0,7668	0,7738	0,7521	<b>0,9466</b>	0,6301
Precision	0,7477	0,7616	0,7736	<b>0,9696</b>	0,7722
Recall	<b>0,9881</b>	0,9706	0,8849	0,9510	0,6427
F1	0,8512	0,8535	0,8255	<b>0,9602</b>	0,7021
AUC	0,6498	0,6667	0,6731	<b>0,9442</b>	0,6836

Źródło: opracowanie własne.

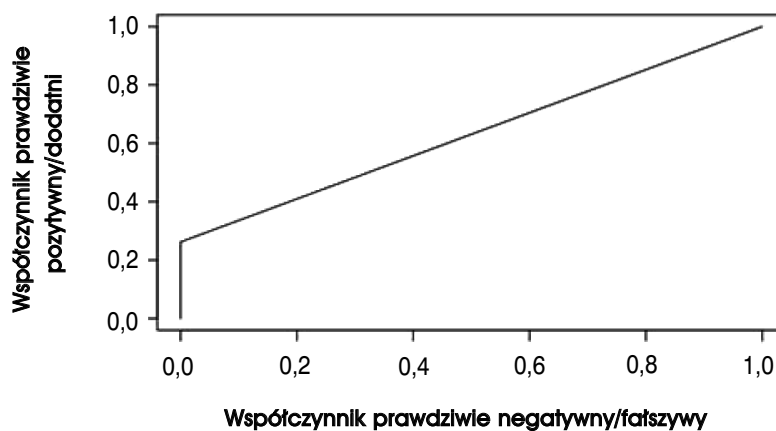
Rysunek 6

Krzywa ROC

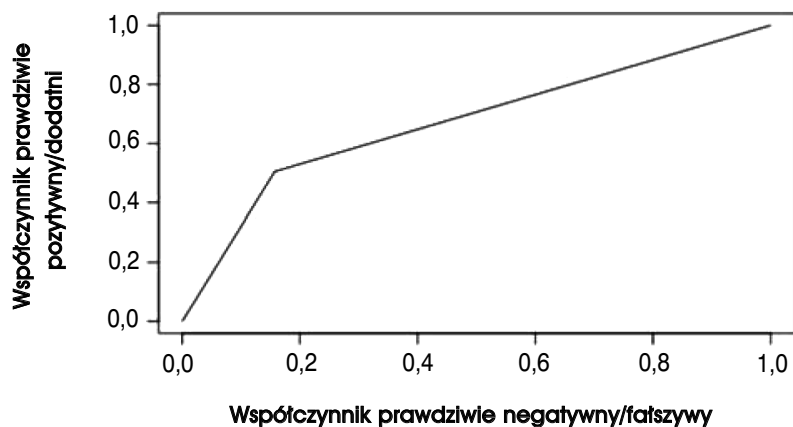
a) maszyna wektorów nośnych



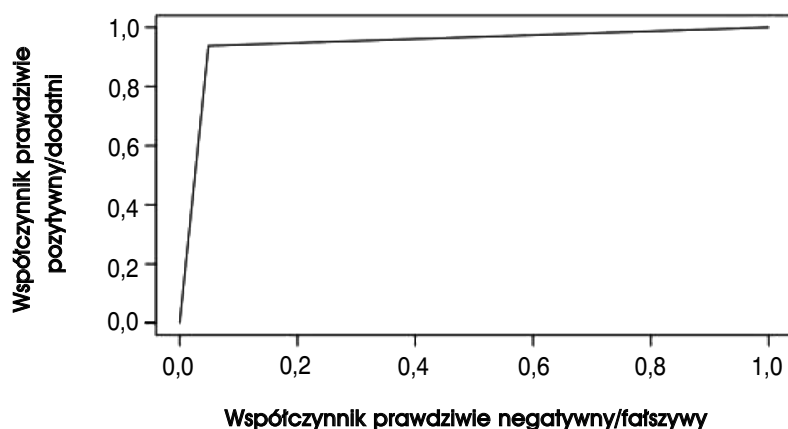
b) drzewo decyzyjne



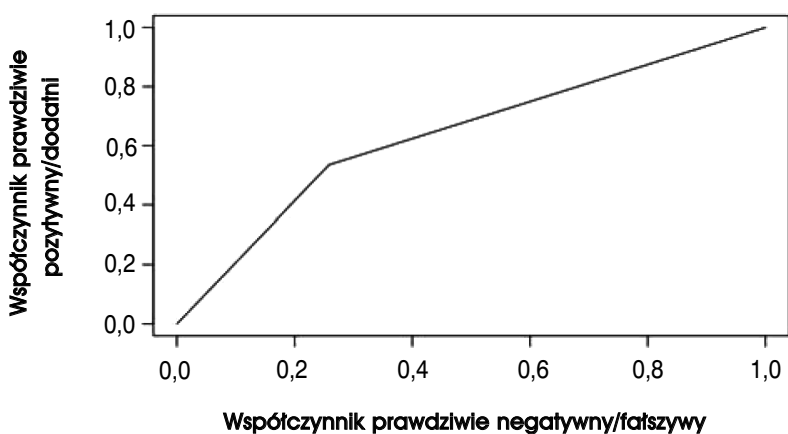
c) naiwny klasyfikator bayesowski



d) AdaBoost



e) regresja logistyczna



Źródło: opracowanie własne.

## Podsumowanie

W czasach logistyki 4.0 analitykę można podzielić m.in. na analitykę predykcyjną, opisową czy nakazową. Analityka predykcyjna ma na celu przewi-

dzenie tego, co wydarzy się w przyszłości, analityka opisowa dostarcza informacji i wyjaśnień na temat tego, co się wydarzyło, natomiast analiza nakazowa dostarcza informacji do podejmowania decyzji (Biazon De Oliveira i in., 2021). Analityka jest często



kojarzona ze sztuczną inteligencją, eksploracją danych i metodami uczenia maszynowego w celu opracowania systemów, które mogą automatycznie wyodrębnić informacje i wzorce w zbiorach danych, aby zapewnić możliwość podejmowania trafnych decyzji, ale także przyczynić się do metamorfozy działania całego łańcucha dostaw w celu poprawy jego bezpieczeństwa czy odporności. Przykładowo przedsiębiorstwo DHL wykorzystuje system Supply Watch (oparty na uczeniu maszynowym oraz analizie), który monitoruje ponad 140 różnych kategorii ryzyka, np. zagrożenia w łańcuchu dostaw takie

jak wąskie gardła czy opóźnienia. System pozwala podjąć szybkie działania w przypadku wykrycia zakłócenia (Schroeder & Lodemann, 2021).

W artykule przedstawiono koncepcję, systemy oraz metody ewaluacji uczenia maszynowego. W ramach eksperymentów zbudowano i oceniono pięć systemów przewidujących ryzyko uszkodzenia towaru. Ustalono, że najlepiej sprawdza się metoda AdaBoost.

W przyszłości autor chciałby się skupić na całym łańcuchu dostaw zamiast na jednym z jego ogniw oraz zastosować inne metody zarówno nadzorowane, jak i nienadzorowane.

## Bibliografia/References

- Anholcer, M. (2017). *Optymalizacja przewozów produktów szybko tracących wartość – modele i algorytmy*. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu.
- Baryannis, G., Dani, S., & Antoniou, G. (2019). Predicting supply chain risks. Using machine learning: The trade-off between performance and interpretability. *Future Generation Computer Systems*, 101, 993–1004. <https://doi.org/10.1016/j.future.2019.07.059>
- Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2018). Supply chain risk management and artificial intelligence: State of the art and future research directions. *International Journal of Production Research*, 57(7), 2179–2022. <https://doi.org/10.1080/00207543.2018.1530476>
- Cunningham, P., Cord, M., & Delany, S. J. (2008). W: M. Cord, & P. Cunningham (Red.), *Machine learning techniques for multimedia. Cognitive technologies* (21–49). Springer. [https://doi.org/10.1007/978-3-540-75171-7\\_2](https://doi.org/10.1007/978-3-540-75171-7_2)
- Biazon de Oliveira, M., Zucchi, G., Lippi, M., Farias, C., Rosa da Silva, N., & Iori, M. (2021). *Lead time forecasting with machine learning techniques for a pharmaceutical supply chain*. Paper presentation. Proceedings of the 23rd International Conference on Enterprise Information Systems on International Conference on Enterprise Information Systems (ICEIS). 26–28.04.2021. <https://www.scitepress.org/Papers/2021/104344/104344.pdf>
- Dridi, S. (2021). *Unsupervised learning – A systematic literature review*. [https://www.researchgate.net/-publication/357380639\\_Unsupervised\\_Learning\\_-\\_A\\_Systematic\\_Literature\\_Review](https://www.researchgate.net/-publication/357380639_Unsupervised_Learning_-_A_Systematic_Literature_Review) (pobrano 22.09.2022).
- Drucker, H., Donghui, W., & Vapnik, V. N. (1999). Support vector machines for spam categorization. *IEEE Transactions on Neural Networks*, 10(5), 1048–1054. <https://doi.org/10.1109/72.788645>
- Iwendi, C., Mahboob, K., Khalid, Z., Javed, A. R., Rizwan, M., & Ghosh, U. (2022). Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multimedia Systems*, 28(4), 1223–1237. <https://doi.org/10.1007/s00-530-021-00774-w>
- Joachimiak, K. (2016). *Zastosowanie technik uczenia maszynowego w przetwarzaniu języka naturalnego* [praca magisterska, Uniwersytet im. Adama Mickiewicza w Poznaniu]. <https://zpjn.wmi.amu.edu.pl/wp-content/uploads/-2019/04/praca-magisterska.pdf>
- Laska, J. (2019). *Przegląd metod uczenia maszynowego stosowanych w analizie wydźwięku* [praca magisterska, Uniwersytet im. Adama Mickiewicza w Poznaniu]. [https://zpjn.wmi.amu.edu.pl/wpcontent/uploads/2019/10/praca\\_magis-terska\\_justyna\\_laska.pdf](https://zpjn.wmi.amu.edu.pl/wpcontent/uploads/2019/10/praca_magis-terska_justyna_laska.pdf)
- Mitchel, T. (1997). *Machine learning*. McGraw-Hill Education.
- Sartias, M. M., & Yasar, A. (2019). Performance analysis of ANN and Naive Bayes classification algorithm for data classification. *International Journal of Intelligent Systems and Applications*, 7, 89–91.
- Schober, P., & Vetter, T. R. (2021). Logistic regression in medical research. *Anesthesia and Analgesia*, 132(2), 365–366. <https://doi.org/10.1213/ANE.0000000000005247>
- Schroeder, M., & Lodemann, S. (2021). A systematic investigation of the integration of machine learning into supply chain risk management. *Logistics*, 5. <https://doi.org/10.3390/logistics5030062>
- Su, J., & Zhang, R. (2006). A fast decision tree learning algorithm. W: *AAAI'06: Proceedings of the 21st national conference on Artificial intelligence*, 1, 500–505.
- Waters, D. (2007). *Supply chain risk management: Vulnerability and resilience*. The Chartered Institute of Logistics and Transportation.
- Wenzel, H., Smit, D., & Sardesai, S. (2019). A literature review on machine learning in supply chain management. *Artificial Intelligence and Digital Transformation in Supply Chain Management: Innovative Approaches for Supply Chains*, 27, 413–441.
- Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion matrix-based feature selection. W: *Proceedings of The 22nd Midwest Artificial Intelligence and Cognitive Science Conference 2011*, 120–127.
- Wyrembek, M. (2022). Wykorzystanie technologii Big Data do predykcji ryzyka opóźnień w łańcuchu dostaw. *Gospodarka Materialowa i Logistyka*, 6, 29–36. <https://doi.org/10.33226/1231-2037.2022.6.4>
- Zhu, J., Zou, H., Rosset, S., & Hastie, T. (2009). Multi-class Adaboost. *Statistics and Its Interface*, 2, 349–360.

### Mgr Mateusz Wyrembek

Doktorant w Szkole Doktorskiej na Uniwersytecie Ekonomicznym w Poznaniu. Specjalista ds. reklamacji w Raben Logistics Polska. Jego zainteresowania naukowe obejmują Big Data i uczenie maszynowe w zarządzaniu łańcuchem dostaw.

### Mgr Mateusz Wyrembek

PhD Student at the Doctoral School at the Poznań University of Economics and Business. Claim specialist at Raben Logistics Polska. His research interests are Big Data and machine learning in supply chain management.