

Wykorzystanie platformy BigML do analizy parametrów geologiczno-górnich kształtujących stan zagrożenia tąpnięciami

MARTA SKIBA 

Instytut Mechaniki Górotworu Polskiej Akademii Nauk, ul. Władysława Reymonta 27, 30-059 Kraków

KINGA ALIBOŹEK

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej, al. Mickiewicza 30, 30-059 Kraków

MATEUSZ ZARĘBA

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, Wydział Geologii, Geofizyki i Ochrony Środowiska, al. Mickiewicza 30, 30-059 Kraków

Streszczenie

Przeprowadzono analizę parametrów geologiczno-górnich, charakteryzujących tąpnięcia zaistniałe w kopalniach Górnośląskiego Zagłębia Węglowego, z wykorzystaniem platformy uczenia maszynowego BigML. W analizach uwzględniono dane odnoszące się do wybranych grup parametrów: strukturalnych, mechanicznych i geometrycznych, charakteryzujących warunki wystąpienia tąpnięć. Zgromadzone dane wykorzystano do budowy modeli uczenia maszynowego do predykcji wartości wybranych parametrów skutków tąpnięć tj. odległości pomiędzy ogniskiem a skutkiem wstrząsu sejsmicznego (1) oraz długości uszkodzeń i/lub zniszczeń wyrobisk górniczych (2). Dla każdego z prognozowanych parametrów opracowano modele bazujące na drzewie decyzyjnym oraz losowym lesie decyzyjnym. Następnie określono skuteczność działania modeli na danych testujących. Dla obu prognozowanych parametrów lepsze rezultaty uzyskano z wykorzystaniem losowego lasu decyzyjnego. Współczynniki determinacji R^2 opracowanych modeli wynosiły 0,782 oraz 0,883 odpowiednio dla parametru odległości pomiędzy ogniskiem a skutkiem wstrząsu oraz długości uszkodzeń i/lub zniszczeń wyrobisk górniczych.

Słowa kluczowe: Platforma BigML; tąpnięcia; uczenie maszynowe; drzewo decyzyjne; losowy las decyzyjny

1. Wstęp

Zagrożenie tąpnięciami jest uwarunkowane czynnikami górniczo-geologicznymi i techniczno-ekonomicznymi. Najistotniejszymi spośród wymienionych są tzw. czynniki naturalne zwane także geologicznymi. Do nich zalicza się pierwotne ciśnienie górotworu, wynikające z dużej głębokości prowadzenia robót górniczych (>760 m), własności skał oraz tektonikę złoża (uskoki). Duży wpływ na zagrożenie tąpnięciami mają również występujące w stropie pokładu tzw. warstwy wstrząsogenne, zwłaszcza o dużej wytrzymałości i grubości.

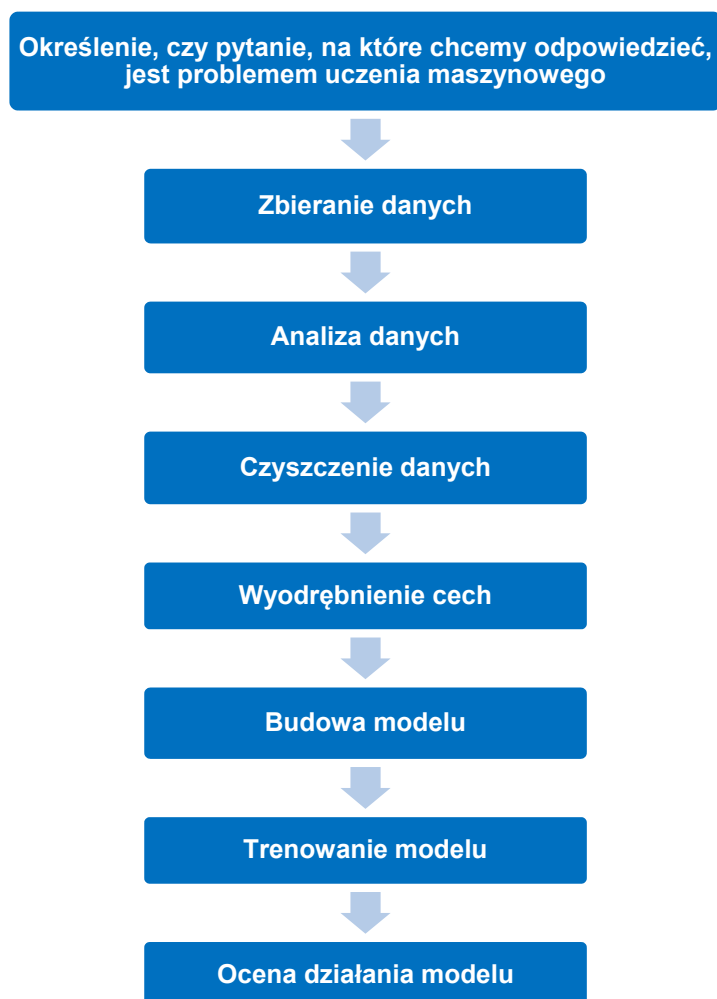
Kolejna grupa to czynniki górnicze, do których zalicza się m.in.: wpływ zaszłości eksploatacji (krawędzie i resztki) w pokładach blisko zalegających oraz skrępowane warunki eksploatacji wynikające z ochrony powierzchni, wybierania resztkowych partii złoża i „niedostatku” robót udostępniających, a czasami również przygotowawczych. W niniejszym artykule przedstawiono analizę wybranych parametrów z miejsc, w których odnotowano 129 tąpnięć. Zestawione dane należą do dwóch: geologicznych i górniczych niezależnych od siebie, grup czynników, kształtujących stan zagrożenia tąpnięciami, a tym samym będących podstawą

uwarunkowań eksploatacji pokładów, w których zaistniało tąpnięcie. Celem prac była kompleksowa analiza tych parametrów z wykorzystaniem platformy uczenia maszynowego BigML. Analiza ta przeprowadzona została na bazie danych źródłowych, dotyczących zaistniałych w przeszłości tąpnięć, pozyskanych z Głównego Instytutu Górnictwa. Uzyskane rezultaty mogą przyczynić się do lepszego rozpoznania zależności pomiędzy parametrami kształtującymi stan zagrożenia tąpnięciami z wykorzystaniem metod uczenia maszynowego. Zaproponowana w pracy metodyka jest nowym podejściem, dotychczas niewykorzystywanym w analizie rozpatrywanego zagadnienia badawczego.

1.1. Platforma BigML – wprowadzenie

W ostatnich latach zaobserwowano znaczny postęp w dziedzinie sztucznej inteligencji, w tym szybki rozwój uczenia maszynowego (ang. *Machine Learning, ML*; Jordan i Mitchell, 2015) i głębokiego uczenia (LeCun i in., 2015). Masowy wzrost mocy obliczeniowej technologii cyfrowej i dostępnych danych przyczyniły się do wzrostu jego wykorzystania w wielu kontekstach (Dwivedi i in., 2021). W związku z tym coraz częściej wdrażane są technologie sztucznej inteligencji, które mogą analizować duże ilości danych, zdobywać umiejętności i wiedzę oraz działać autonomicznie (Simsek i in., 2019). Uczenie maszynowe zostało zdefiniowane jako „szeroki podzbiór sztucznej inteligencji, w którym program komputerowy stosuje algorytmy i modele statystyczne do konstruowania złożonych wzorców wnioskowania na podstawie danych” (Leavitt i in., 2021).

Jednym z kluczowych wyzwań związanych z pełnym wykorzystaniem uczenia maszynowego jest konieczność posiadania doświadczenia w kodowaniu (lub programowaniu) oraz nabycia infrastruktury związanej z uczeniem maszynowym. Bariery te można pokonać dzięki dostępności różnych platform, które zapewniają zautomatyzowane i wolne od kodowania usługi ML, a także infrastrukturę ML (w postaci



Rys. 1. Schemat postępowania podczas realizacji Projektu uczenia maszynowego z wykorzystaniem platformy BigML

usługi w chmurze lub oprogramowania) (Sundberg i Holmström, 2023). Jedną z takich platform uczenia maszynowego jest platforma BigML.

Tradycyjne podejście do uczenia maszynowego polega na gromadzeniu danych odnoszących się do zestawu zmiennych (zwanymi również cechami) i obserwacji (wyników). Są one następnie przetwarzane i wprowadzane do algorytmu ML za pomocą kodowania (tj. programowania). Skrypt kodu zawiera procedurę obsługi danych wejściowych, która jest wykonywana w celu szkolenia i weryfikacji wydajności algorytmu. Na koniec analizowane są wyniki analizy. W przypadku platform uczenia maszynowego należy postępować zgodnie z algorytmem zaprezentowanym na rysunku 1.

Platforma BigML została utworzona w 2011 roku i ma siedzibę w Corvallis w USA. Celem BigML było stworzenie usługi uczenia maszynowego łatwej w użyciu i bezproblemowej w integracji. W BigML wykorzystano drzewo decyzyjne jako jedno z narzędzi uczenia maszynowego do analizy zbioru danych. Dzieje się tak, ponieważ drzewo decyzyjne jest narzędziem wspomagającym podejmowanie decyzji, które można postrzegać jako „drzewiasty” wykres decyzji w wynikach. Ten algorytm jest zwykle użyteczny w badaniach operacyjnych, w szczególności w analizie decyzji, aby pomóc w znalezieniu najlepszego rozwiązania do osiągnięcia celu (BigML, 2023). Platforma ta ma silny program akademicki, który wspiera edukację studentów oraz badania naukowe. W przeciwieństwie do innych platform, BigML pozwala użytkownikom eksportować utworzone modele ML w różnych językach, tym samym podkreślając pojęcie przenośnego ML. Platforma ta obsługuje przede wszystkim dane tabelaryczne.

Coraz częściej platforma BigML wykorzystywana jest także do zagadnień prognozy. W pracy (Zainudin i Shamsuddin, 2016) autorzy wykorzystali platformę BigML do przewidywania najpopularniejszego miejsca występowania dengi, aby uzyskać możliwość wczesnego ostrzeżenia i uświadamiania ludzi. BigML wykorzystuje algorytmy drzewa decyzyjnego do analizy danych i przewidywania miejsca występowania dengi w Malezji. Ponadto uczenie maszynowe okazało się szczególnie przydatne w kontekstach organizacyjnych (Sturm i in., 2021).

2. Parametryzacja przyczyn i skutków tąpnięć

Ilościowe kryteria kwalifikowania skłonności górotworu do tąpnięć, podane w pracy Konopko (Konopko, 2006) określone są w oparciu o równoważnik wytrzymałościowy stropu piaskowcowego, aktywność sejsmiczną oraz charakterystykę wytrzymałościowo-deformacyjną układu strop-pokład-spąg. W pokładzie i górotworze, znajdujących się w warunkach niezaburzonych tektonicznie, o zagrożeniu tąpnięciami stanowią: stan naprężenia, właściwości mechaniczne pokładu i skał otaczających go oraz sejsmiczność generowana robotami górnymi (Goszcz, 1988). Ta ostatnia jest rezultatem zachodzącej destrukcji podbieranych warstw skalnych o wysokich parametrach wytrzymałościowych, a w szczególności tzw. warstw wstrząsogennych. Stanowią je grube, monolityczne warstwy piaskowców i/lub iłowców, o odpowiednio wysokiej wytrzymałości i grubości. Zagrożenie tąpnięciami zwiększa się wraz z głębokością prowadzenia robót (Konopko, 1994) i wynikającym stąd zwiększonym stanem naprężenia w górotworze i gradientem jego zmian. Przy ogólnym wysokim naprężeniu, tąpnięcia występują przy relatywnie niskich energiach wstrząsów górotworu (Dubiński i Konopko, 2000; Konopko i Patyńska, 2008).

Do analiz z wykorzystaniem platformy BigML wytypowano 10 parametrów przyczyn tąpnięć: wydobyte kopalń tąpniących [mln ton] (1), minimalna grubość pokładu [m] (2), minimalny kąt upadu pokładu [°] (3), wysokość wyrobiska/eksploatacji [m] (4), grubość warstwy tąpniącej [m] (5), głębokość pokładu [m] (6), wytrzymałość węgla na ściskanie (R_c) [MPa] (7), odległość spągu warstwy tąpniącej do stropu pokładu [m] (8), maksymalny zrzut uskoku lokalnego [m] (9) oraz minimalny zrzut uskoku lokalnego [m] (10). Parametry te wykorzystano do budowy modeli do prognozy dwóch parametrów tąpnięć tj. odległości pomiędzy ogniskiem a skutkiem wstrząsu sejsmicznego [m] oraz długości uszkodzeń i/lub zniszczeń wyrobisk górniczych [m] (Patyńska, 2011).

3. Analizy z wykorzystaniem platformy BigML

Analizy dotyczące opracowania optymalnych modeli dla rozpatrywanego zadania predykcji wybranych parametrów skutków tąpnięć tj. odległości pomiędzy ogniskiem a skutkiem wstrząsu sejsmicznego oraz sumarycznej długości uszkodzeń i/lub zniszczeń wyrobisk górniczych przeprowadzono z wykorzystaniem platformy BigML.

Z dostępnego zbioru danych, 80% (103 elementy) wykorzystano do treningu wybranego modelu, natomiast pozostałe 20% (26 elementów) stanowił zbiór testowy, na podstawie którego dokonywano ostatecznej oceny działania opracowanego modelu. Dla każdego z prognozowanych parametrów opracowano modele bazujące na drzewach decyzyjnych oraz zespole drzew decyzyjnych (tzw. losowy las decyzyjny). Następnie określono skuteczność działania modeli na danych testujących.

Platforma BigML jako miarę dopasowania danego modelu do danych stosuje wartość współczynnika determinacji (R^2) oraz średni błąd bezwzględny (ang. *Mean Absolute Error*, MAE), wyznaczone dla zbioru testowego. Współczynnik determinacji identyfikuje odsetek wariancji w zmiennej przewidywanej, który jest wyjaśniony przez zmienną wejściową lub zmienne wejściowe. Oblicza się go ze wzoru:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

gdzie:

- \hat{y} – wartość przewidywana zmiennej zależnej,
- \bar{y} – średnia wartość zmiennej zależnej y ,
- n – liczba elementów zbioru testowego.

Kolejną miarą jest średni błąd bezwzględny MAE. Informuje on ile średnio w okresie prognoz, będzie wynosić odchylenie od wartości rzeczywistej. Oblicza się go ze wzoru:

$$MAE = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n} \quad (2)$$

gdzie:

- \hat{y} – wartość przewidywana zmiennej zależnej,
- y – wartość obserwowana zmiennej zależnej,
- n – liczba elementów zbioru testowego.

4. Prognoza za pomocą drzew decyzyjnych oraz losowego lasu decyzyjnego

Jednym z najważniejszych zadań w procesie prognozy jest wybór i przygotowanie danych do trenowania systemu. Należy wybrać wysokiej jakości i dobrze ustrukturyzowane dane. W przeciwnym wypadku wprowadzimy do systemu szum, który będzie generował prognozy o niewielkiej wartości. Chociaż podczas procesu uczenia algorytmu system może wykryć i odrzucić dane, które są uważane za zbędne, ważne jest, aby dokonać wyboru danych, które cechują się wysoką informatywnością, z punktu widzenia przygotowywanej prognozy.

Pierwszym krokiem podczas analiz z wykorzystaniem platformy BigML było utworzenie źródła danych. Stanowi ono dane nieprzetworzone, z których następnie tworzy się zestawy danych. Do analiz wytypowano wyłącznie dane numeryczne, a prognozowany parametr był ostatnim atrybutem (stanowi ostatnią kolumnę pliku źródłowego CSV).

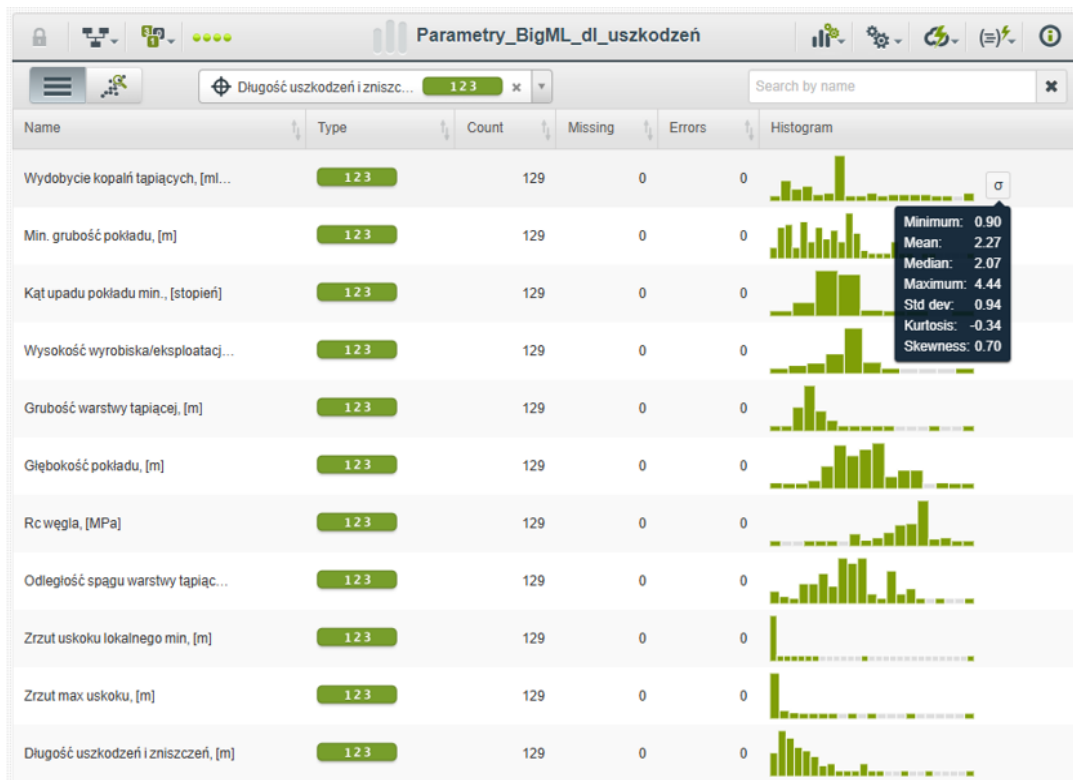
Następnie z pliku źródłowego utworzono zestaw danych (funkcja „1-Click Dataset”), służący do wstępnej analizy parametrów a następnie do budowy modelu predykcyjnego. Otrzymany zestaw danych, uwzględniający rozkład poszczególnych parametrów oraz ich podstawowe miary statystyczne zaprezentowano na rysunku 2.

Kolejnym krokiem był podział otrzymanego zbioru danych na dane trenujące (103 przykłady, 80% danych) oraz dane testowe (26 przykładów, 20% danych) za pomocą funkcji „Training | Test Split”. Pierwszy z utworzonych zbiorów wykorzystano do budowy modelu predykcyjnego, natomiast drugi, zawierający przykłady, których nie użyto do szkolenia modelu, posłużył do oceny działania opracowanego modelu.

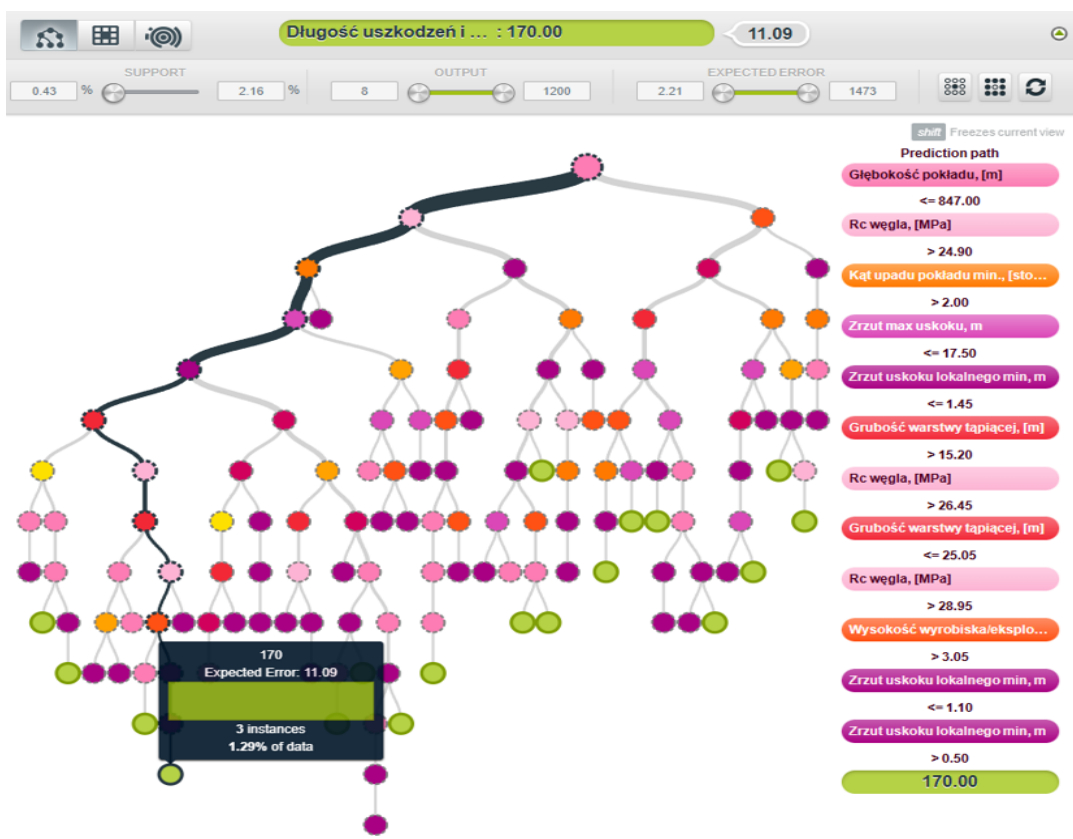
Następnie utworzono model bazujący na drzewie decyzyjnym dla pierwszego z prognozowanych parametrów tj. długości uszkodzeń i/lub zniszczeń wyrobisk górniczych (Rys. 3).

Drzewo decyzyjne to jedna z najczęściej wykorzystywanych technik analizy danych. Drzewo składa się z korzenia oraz gałęzi prowadzących z korzenia do kolejnych wierzchołków. Wierzchołki, z których

wychodzi co najmniej jedna krawędź, są nazywane węzłami, a pozostałe wierzchołki – liśćmi. W każdym węźle sprawdzany jest pewien warunek dotyczący danej obserwacji, i na jego podstawie wybierana jest jedna z gałęzi prowadząca do kolejnego wierzchołka.



Rys. 2. Zestaw danych BigML do predykcji wybranych parametrów skutków tąpnięć



Rys. 3. Drzewo decyzyjne do prognozy wartości długości uszkodzeń i/lub zniszczeń wyrobisk górniczych

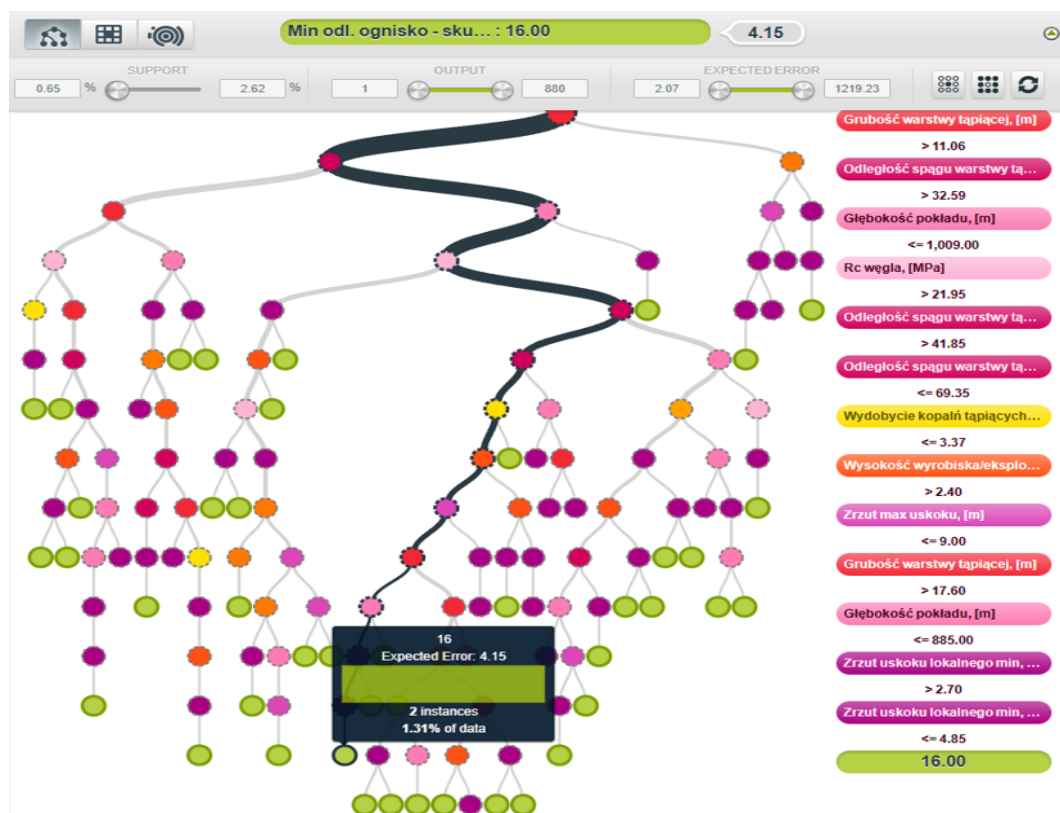
Dokonano oceny działania modelu na danych testujących (26 przykładów), uwzględniając 10-krotny proces losowania ciągu uczącego. Uzyskane rezultaty przedstawiono w tabeli 1.

Tab. 1. Wartości miar dopasowania modelu do danych dla prognozy długości uszkodzeń i/lub zniszczeń wyrobisk górniczych

Lp.	R^2	MAE [m]
1	0,79	56,99
2	0,72	71,34
3	0,88	50,98
4	0,70	63,17
5	0,86	52,35
6	0,74	56,42
7	0,70	64,15
8	0,74	57,75
9	0,83	59,34
10	0,88	48,63
Średnia	0,78	58,11
Odchylenie standardowe	0,073	6,805

Analiza tabeli 1 wskazuje, że opracowany model bazujący na drzewie decyzyjnym posiada dobre dopasowanie do danych rzeczywistych. W wyniku analiz dla 10 losowań ciągu uczącego uzyskano średnią wartość współczynnika determinacji na poziomie 0,78, natomiast średni błąd bezwzględny wyniósł 58,11 m. Należy jednak zauważyć, że nawet w przypadku najlepszego dopasowania ($R^2 = 0,88$), wartość MAE była wysoka (49 m-51 m).

Analogiczne analizy przeprowadzono dla drugiego z prognozowanych parametrów – odległości pomiędzy ogniskiem a skutkiem wstrząsu sejsmicznego. Model drzewa decyzyjnego do prognozy tego parametru przy użyciu platformy BigML przedstawiono na rysunku 4.



Rys. 4. Drzewo decyzyjne do prognozy wartości minimalnej odległości pomiędzy ogniskiem a skutkiem wstrząsu sejsmicznego

Następnie dokonano oceny działania modelu na danych testujących (26 przykładów), uwzględniając 10-krotny proces losowania ciągu uczącego. Uzyskane rezultaty przedstawiono w tabeli 2.

Tab. 2. Wartości miar dopasowania modelu do danych dla prognozy minimalnej odległości pomiędzy ogniskiem a skutkiem wstrząsu sejsmicznego

Lp.	R^2	MAE [m]
1	0,71	48,65
2	0,75	50,72
3	0,77	40,98
4	0,67	29,77
5	0,64	56,46
6	0,62	65,30
7	0,67	36,98
8	0,74	49,00
9	0,69	44,28
10	0,76	41,85
Średnia	0,70	46,40
Odchylenie standardowe	0,052	10,049

W przypadku prognozy odległości pomiędzy ogniskiem a skutkiem wstrząsu sejsmicznego za pomocą drzewa decyzyjnego uzyskano nieco gorsze wyniki dopasowania modelu do danych rzeczywistych (średnia wartość współczynnika determinacji $R^2 = 0,70$). Dla 10 losowań ciągu uczącego średni błąd bezwzględny MAE wyniósł 46,40 m. Dla najlepszego dopasowania ($R^2 = 0,77$) wartość ta była na poziomie 41 m.

Kolejnym etapem prac była prognoza wybranych parametrów skutków tąpnięć z wykorzystaniem grupy drzew decyzyjnych tzw. losowego lasu decyzyjnego. Algorytm losowego lasu decyzyjnego (ang. random forest) stanowi jeden z najciekawszych oraz najbardziej użytecznych algorytmów uczenia maszynowego. Idea lasu losowego opiera się na stworzeniu z góry zdefiniowanej liczby drzew decyzyjnych, które są budowane na podstawie konkretnej liczby losowych obserwacji ze zbioru treningowego, opisanych zadaną liczbą zmiennych objaśniających. W przypadku zadań regresji (prognozy wartości zmiennej ciągłej) jako wynik działania uzyskujemy przewidywaną średnią poszczególnych drzew. Metoda losowego lasu decyzyjnego jest sposobem na uniknięcie nadmiernego dopasowania modelu do zestawu treningowego – co stanowi częstą przypadłość mocno rozbudowanych pojedynczych drzew decyzyjnych. Ponadto do najważniejszych zalet lasu losowego należy odporność na brak danych, dużą liczbę zmiennych objaśniających czy wartości odstające oraz możliwość dokładniejszego odtworzenia złożonych zależności w porównaniu do pojedynczych drzew decyzyjnych.

W przypadku analiz opisywanych w niniejszej pracy użyto grypy 5 drzew decyzyjnych, generując odpowiednie modele lasu losowego dla każdego z prognozowanych parametrów. Do budowy modeli wykorzystano zbiory treningowe, składające się z 80% dostępnych danych tj. 103 przykłady uczące. Tak utworzony losowy las decyzyjny generował wynik, będący uśrednioną wartością prognozy wszystkich rozpatrywanych drzew decyzyjnych.

Następnie dokonano oceny działania utworzonych modeli na danych testowych (26 przykładów) dla 10 losowań ciągu uczącego. Uzyskane wyniki prognozy długości uszkodzeń i/lub zniszczeń wyrobisk górniczych oraz minimalnej odległości pomiędzy ogniskiem a skutkiem wstrząsu sejsmicznego zaprezentowano odpowiednio w tabeli 3 oraz 4.

Tab. 3. Wartości miar dopasowania modelu lasu losowego do danych dla prognozy długości uszkodzeń i/lub zniszczeń wyrobisk górniczych

Lp.	R^2	MAE [m]
1	0,97	10,05
2	0,93	15,79
3	0,97	9,55
4	0,92	19,38
5	0,96	11,96
6	0,79	25,00
7	0,79	30,74
8	0,80	29,42
9	0,74	48,16
10	0,96	16,12
Średnia	0,88	21,62
Odchylenie standardowe	0,091	12,027

Tab. 4. Wartości miar dopasowania modelu lasu losowego do danych dla prognozy minimalnej odległości pomiędzy ogniskiem a skutkiem wstrząsu sejsmicznego

Lp.	R^2	MAE [m]
1	0,72	22,10
2	0,83	10,71
3	0,80	13,83
4	0,82	24,18
5	0,74	15,95
6	0,77	35,77
7	0,71	38,82
8	0,81	6,62
9	0,82	22,10
10	0,80	19,38
Średnia	0,78	20,95
Odchylenie standardowe	0,044	10,222

W przypadku obu prognozowanych parametrów skutków tąpnięć z wykorzystaniem algorytmu losowego lasu decyzyjnego uzyskano poprawę wyników dopasowania modeli do danych rzeczywistych (patrz tabela 1 i 3 oraz tabela 2 i 4). Dla parametru długości uszkodzeń i/lub zniszczeń wyrobisk górniczych średnia wartość współczynnika determinacji R^2 wzrosła o 12,8%, natomiast dla drugiego parametru, tj. minimalnej odległości pomiędzy ogniskiem a skutkiem wstrząsu sejsmicznego wzrost ten wyniósł 11,4%. Dla najlepszego dopasowania wartość współczynnika determinacji była na poziomie 0,97 dla długości uszkodzeń i/lub zniszczeń wyrobisk górniczych oraz 0,83 dla minimalnej odległości pomiędzy ogniskiem a skutkiem wstrząsu sejsmicznego. Poprawie uległy także średnie wartości średniego błędu bezwzględnego prognozy. Dla parametru długości uszkodzeń i/lub zniszczeń wyrobisk górniczych średnia wartość MAE spadła o 62,8%, natomiast dla drugiego z prognozowanych parametrów spadek ten wyniósł 54,8%. Uzyskane rezultaty świadczą o dobrych zdolnościach predykcyjnych modeli bazujących na losowym lesie decyzyjnym.

5. Wnioski

W artykule przeprowadzono analizę parametrów geologiczno-górniczych, charakteryzujących tąpnięcia zaistniałe w kopalniach Górnośląskiego Zagłębia Węglowego, z wykorzystaniem platformy uczenia maszynowego BigML. Z parametrów tych wyselekcjonowano „mierzałe” i najbardziej kształtujące stan zagrożenia tąpnięciami parametry (parametry przyczyn) oraz parametry pokazujące skalę ich oddziaływania (parametry skutków). Dla każdego z prognozowanych parametrów skutków tąpnięć opracowano modele bazujące na drzewach decyzyjnych oraz losowych lasach decyzyjnych. Dla obu prognozowanych parametrów lepsze rezultaty uzyskano z wykorzystaniem losowego lasu decyzyjnego. Współczynniki determinacji R^2 opracowanych modeli wynosiły 0,782 oraz 0,883 odpowiednio dla parametru odległości pomiędzy ogniskiem a skutkiem wstrząsu oraz sumarycznej długości uszkodzeń i/lub zniszczeń wyrobisk górniczych, co stanowiło wzrost o 12,8% oraz 11,4% w stosunku do średniej wartości współczynników determinacji uzyskanych dla modeli bazujących na drzewie decyzyjnym. Dla najlepszego dopasowania wartość współczynnika determinacji była na poziomie 0,97 dla długości uszkodzeń i/lub zniszczeń wyrobisk górniczych oraz 0,83 dla minimalnej odległości pomiędzy ogniskiem a skutkiem wstrząsu sejsmicznego.

Praca została wykonana w ramach prac statutowych realizowanych w IMG PAN w roku 2024, finansowanych przez Ministerstwo Nauki i Szkolnictwa Wyższego oraz Projektu badawczego nr 5 (FBW/D/2023-2024/05/2024), finansowanego z Funduszu Badań Własnych IMG PAN.

Literatura

- [1] BigML, (2023): *About BigML*. com. <https://bigml.com/>
- [2] Dubiński J., Konopko W., (2000): *Tąpnięcia-ocena-prognoz-zwalczanie*. Wydawnictwo GIG, Katowice.
- [3] Dwivedi Y.K., Hughes L., Ismagilova E., Aarts G., Coombs C., Crick T., Duan Y., Dwivedi R., Edwards J., Eirug A.,

- Galanos V., Ilavarasan P.V., Janssen M., Jones P., Kar A.K., Kizgin H., Kronemann B., Lal B., Lucini B., ... Williams M.D. (2021): *Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice, and policy*. International Journal of Information Management, 57, 101994.
- [4] Goszcz A., (1988): *Wpływ niektórych czynników technologicznych na stan zagrożenia wstrząsami górnictwami i tapaniami*. Publikacje Instytutu Geofizyki Polskiej Akademii Nauk M-10 (213), s. 141-153.
- [5] Jordan M.I., Mitchell T.M., (2015): *Machine learning: Trends, perspectives, and prospects*. Science, 349 (6245), 255-260.
- [6] Konopko W., (2006): *Wytobiska–nie pokłady–zagrożone tapaniami. Podstawy nowelizacji przepisów*. Bezpieczeństwo Pracy i Ochrony Środowiska w Górnictwie. Nr 8.
- [7] Konopko W., Patyńska R., (2008): *Warunki występowania tapanieć w kopalniach węgla kamiennego*. Przegląd Górniczy nr 1, s. 12–17.
- [8] Leavitt K., Schabram K., Hariharan P., Barnes C.M., (2021): *Ghost in the machine: On organizational theory in the age of machine learning*. Academy of Management Review, 46 (4), 750-777.
- [9] LeCun Y., Bengio Y., Hinton G., (2015). *Deep learning*. Nature, 521 (7553), 436-444.
- [10] Simsek Z., Vaara E., Paruchuri S., Nadkarni S., Shaw J.D., (2019): *New ways of seeing big data*. Academy of Management Journal, 62(4), 971-978.
- [11] Patyńska R., (2011): *Strefy zagrożenia tapaniami dla poszczególnych stopni zaliczeń*. Wiadomości Górnicze nr 1, s. 7-13.
- [12] Sturm T., Gerlach J.P., Pumplun L., Mesbah N., Peters F., Tauchert C., Nan N., Buxmann P., (2021): *Coordinating human and machine learning for effective organizational learning*. MIS Quarterly, 45 (3b), 1581-1602.
- [13] Sundberg L., Holmström J. (2022): *Towards 'lightweight' artificial intelligence: a typology of AI service platforms*. In Proceedings of the American Conference on Information Systems. Minneapolis, MN: AMCIS.
- [14] Zainudin Z., Shamsuddin S.M., (2016): *Predictive analytics in Malaysian dengue data from 2010 until 2015 using BigML*. 8. 18-30.

The use of the BigML platform for the analysis of geological-mining parameters influencing the rockbursts hazard

Abstract

An analysis of geological-mining parameters, characterizing rockbursts that occurred in the mines of the Upper Silesian Coal Basin, was carried out using the BigML machine learning platform. The analyses included data relating to selected groups of parameters: structural, mechanical and geometric, characterizing the conditions of occurrence of rockburst hazard. The collected were used to build machine learning models for predicting the values of selected parameters of the effects of rockbursts, i.e. distance between the focus and the effect of a seismic tremor (1) and the length of damage and/or destruction of the excavations (2). Models based on a decision tree and a random decision forest were developed for each of the predicted parameters. The performance of the models was then determined on test data. For both forecast parameters, better results were obtained using the random decision forest. The R^2 coefficients of determination of the developed models were 0.782 and 0.883 for the (1) and (2) parameter respectively.

Keywords: BigML platform; rockbursts; machine learning; decision tree; random decision forest