

## Rozdział 11

# Przykłady zastosowania systemu DISESOR w analizie i predykcji zagrożeń

Andrzej Janusz, Marek Grzegorowski, Michał Kozielski, Zdzisław Krzystanek, Marek Sikora, Dominik Ślęzak, Łukasz Wróbel

---

### 11.1 Wstęp

Istotnym wyzwaniem dla górnictwa węgla kamiennego jest zapewnienie bezpieczeństwa pracującym pod ziemią ludziom i maszynom. W górnictwie węgla kamiennego występuje wiele zagrożeń naturalnych, lecz dwa najważniejsze z nich to: zagrożenie metanowe i sejsmiczne. W niniejszym rozdziale zaprezentowano zastosowanie systemu DISESOR do predykcji każdego z wymienionych zagrożeń.

Zadanie predykcji metanu w wyrobisku, w którym prowadzona jest eksploatacja górnicza, jest zadaniem niezwykle trudnym ze względu na wiele czynników mających wpływ na rejestrowane na wylocie ściany stężenia metanu. Wylot ściany jest miejscem, w którym rejestrowane są najwyższe stężenia metanu, dlatego właśnie w tym miejscu zgodnie z obowiązującymi przepisami umieszczane są metanomierze odpowiedzialne za automatyczne (profilaktyczne) wyłączenia energii elektrycznej. Modele matematyczne łączące wydzielanie metanu z teoretyczną metanością pokładu, sposobem przewietrzania i własnościami geologicznymi wyrobiska [11.16, 11.19, 11.25] pozwalają w sposób ogólny prognozować, z jakimi średnimi stężeniami metanu będzie wiązało się prowadzenie eksploatacji w danym wyrobisku. Modele takie nie nadają się jednak do bezpośredniej krótko i średnioterminowej predykcji stężenia metanu. Metody podejmujące próby prognozy zagrożenia metanowego, rozumianego jako prognoza konkretnej wartości stężenia metanu albo ocena ryzyka związanego z wybuchem gazu, wykorzystują zazwyczaj techniki statystyczne [11.3]. Dla kilkuminutowej predykcji stężenia metanu stosowano także techniki maszynowego uczenia [11.22, 11.23].

Ustalenie aktualnego stopnia tzw. zagrożenia sejsmicznego (w szczególności zagrożenia destrukcyjnym wstrząsem wysokoenergetycznym zwanym tąpnię-

ciem) w miejscach prowadzenia robót jest jednym z głównych zadań stacji geofizycznych w kopalniach węgla kamiennego. Wstrząsy mogą stanowić poważne zagrożenia dla życia załóg górniczych oraz mogą być przyczyną niszczenia wyrobisk i znajdującego się w nich sprzętu. Coraz lepsze sejsmiczne i sejsmoakustyczne systemy monitorowania pozwalają na lepsze zrozumienie procesów zachodzących w górotworze [11.4], prowadzą także do definiowania lepszych metod przewidywania zagrożeń sejsmicznych [11.5]. Dokładność tworzonych metod daleka jest jednak od doskonałości. Nowe metody oceny zagrożenia sejsmicznego to m.in.: metoda probabilistyczna [11.17] przewidująca wartość energii wstrząsów sejsmicznych jakie zarejestrowane zostaną w określonym horyzoncie prognozy, metoda prognozy liniowej [11.15] przewidująca wartość energii umownej (sejsmicznej i sejsmoakustycznej) jaka wydzieli się w określonym horyzoncie prognozy lub metoda tzw. funkcji wskaźnikowych, która próbuje estymować prawdopodobieństwo wystąpienia silnego wstrząsu [11.2]. Złożoność procesów powstawiania wstrząsów sejsmicznych, a także duża dysproporcja pomiędzy liczbą rejestrowanych zjawisk sejsmicznych o niskiej energii a zjawiskami o energiach większych (np.  $> 1 * 10^4 \text{J}$ ) powoduje, że techniki statystyczne w przewidywaniu zagrożeń sejsmicznych okazują się niewystarczające, dlatego coraz bardziej popularne staje się wykorzystanie do tego celu metod inteligencji obliczeniowej i maszynowego uczenia.

## 11.2 Metody oceny zagrożenia

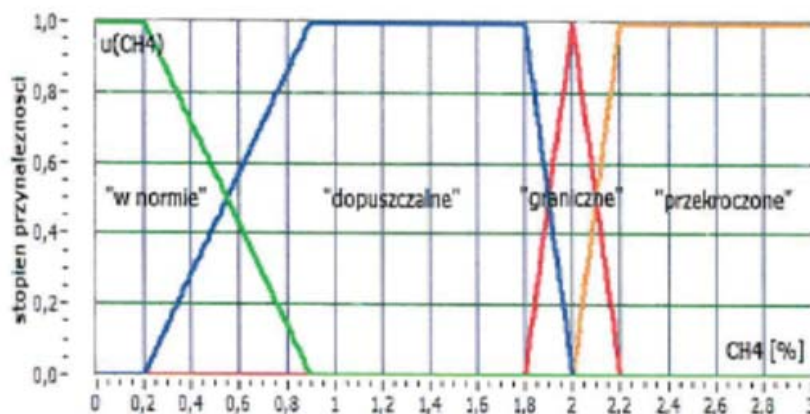
Standardowe metody oceny zagrożeń naturalnych, określane jako eksperckie, dotyczą nie prognoz, lecz oceny stanu aktualnego. Metod takich jest wiele, a w niniejszym punkcie zostaną ogólnie omówione: metoda oceny zagrożenia metanowego bazująca na wywiadach z dyspozytorami metanowymi oraz metoda kompleksowa (dla zagrożeń sejsmicznych).

### 11.2.1 Metoda oceny zagrożenia metanowego

Zagrożenie metanowe w kopalniach węgla kamiennego należy do najczęściej spotykanych i związane jest z nim niebezpieczeństwo zapoczątkowania wybuchu metanu lub pyłu węglowego. Metan jest więc jednym z najczęściej monitorowanych składników atmosfery kopalnianej. Dla metanu, zgodnie z przepisami, próg alarmowy definiuje się od 1 do 2%  $\text{CH}_4$  (zależnie od miejsca zabudowy), natomiast próg ostrzegawczy (tzw. margines bezpieczeństwa) ustala dział wentylacji. Do zabezpieczenia metanowego w typowych rejonach wydobywczych służą metanomierze wyłączająco-rejestrujące. Lokalizacja metanomierzy wyłączających energię elektryczną to umowne punkty w wyrobisku, w pobliżu występujących tam urządzeń elektrycznych tzw. przyścianowych pociągów aparaturowych, gdyż nie można przewidzieć miejsca wypływu metanu do wyrobiska.

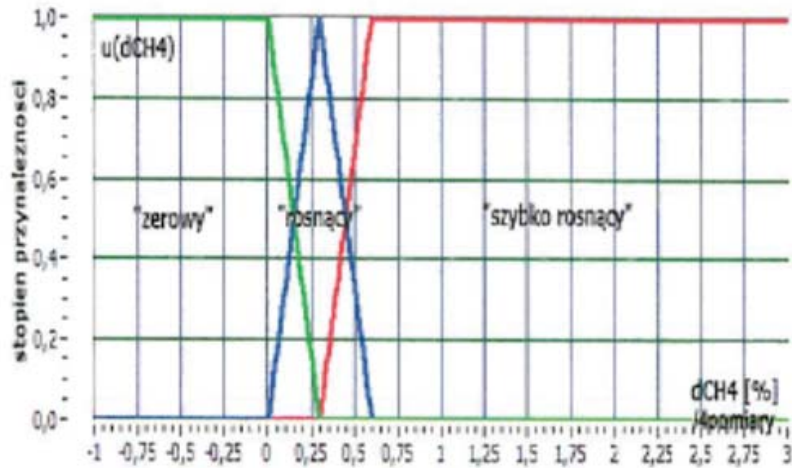
W celu wspomagania dyspozytora przy kontroli metanu możliwe jest wykorzystanie algorytmu [11.6], w którym zaimplementowano schemat wnioskowania Mamdaniego z dwiema metodami wyostrzania: COG (ang. Center of Gravity)

oraz MOM (ang. Medium of Maximum) [11.18]. W ramach prezentowanej w tym rozdziale metody została opracowana baza reguł oraz funkcje przynależności dla metanomierzy rejestrująco-wyłączających o przepisowych progach wyłączeń  $2\%CH_4$ , które instalowane są w rejonach wydobywczych ścian. Został również zaproponowany podział wartości stężenia metanu na podstawie informacji opisowej, jaką posługuje się dyspozytor monitorujący takie zagrożenie. W przypadku przedstawianej metody wartością "graniczną" jest "około  $2\%CH_4$ ". Algorytm rozmyty wnioskuje znacznie szybciej, niż czyni to dyspozytor, co jest znaczącą zaletą przy bardzo szybkich zmianach stężenia gazu. Informacją wejściową dla systemu jest przyrost stężenia metanu w czasie, dlatego system szybko reaguje na wzrosty stężenia metanu np. odcinając energię elektryczną, a przez to zabezpiecza przed potencjalnym wybuchem wstrzymując pracę urządzeń, co nie dopuszcza do zagrażających wybuchem wartości stężenia tego gazu. W przypadku nagłych zmian stężenia oraz zbyt częstego wyłączania energii elektrycznej, a takie sytuacje mają miejsce, próg alarmowy oparty na samych przyrostach metanu nie zda egzaminu. Dlatego w algorytmie wprowadzono także aktualną wartość stężenia metanu. Funkcje przynależności zdefiniowane dla zmiennych wejściowych, które zostały przedstawione na rysunkach 11.1 oraz 11.2, są efektem wywiadów z kilkunastoma dyspozytorami metanometrii. Powyższa metoda została zaimplementowana w systemie DISESOR, co ilustrują rysunki 11.3 oraz 11.4 przedstawiające zmienne wejściowe, ich zakresy oraz podziały.



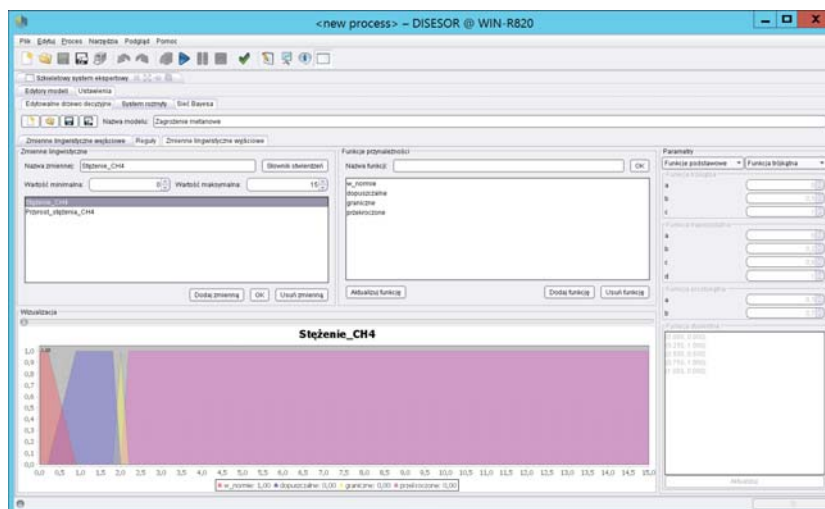
Rys. 11.1. Podział zmiennej wejściowej "przyrost stężenia  $CH_4$ " na zbiory rozmyte

W tabeli 11.1 przedstawiono bazę reguł dyspozytora, która została zaimplementowana w systemie DISESOR. Zmienne "stężenie  $CH_4$ " oraz "przyrost stężenia  $CH_4$ " to przesłanki reguł (patrz rysunki 11.1 i 11.2), natomiast "stopień zagrożenia" to konkluzja reguł rozmytych.

Rys. 11.2. Podział zmiennej wejściowej "stężenia  $CH_4$ " na zbiory rozmyteRys. 11.3. Podział zmiennej wejściowej "przyrost stężenia  $CH_4$ " na zbiory rozmyte - realizacja w środowisku systemu DISESOR (moduł System Ekspercki)

### 11.2.2 Metoda oceny zagrożenia sejsmicznego

Metoda kompleksowa dla zagrożeń sejsmicznych [11.1] bazuje na wynikach ocen innych metod eksperckich, które z kolei bazują już na danych pomiarowych. W skład metody kompleksowej wchodzi metody szczegółowe, właściwe dla określania stanu zagrożenia tąpnięciami pokładowymi i stropowymi oraz metoda oceny potencjalnego stanu zagrożenia tąpnięciami, bazująca na uogólnionej interpretacji geologiczno-górnich przyczyn występowania tąpnięć w kopalniach



Rys. 11.4. Podział zmiennej wejściowej "stężenia  $CH_4$ " na zbiory rozmyte - realizacja w środowisku systemu DISESOR (moduł System Ekspertyki)

Tab. 11.1. Baza reguł dla czujników metanu

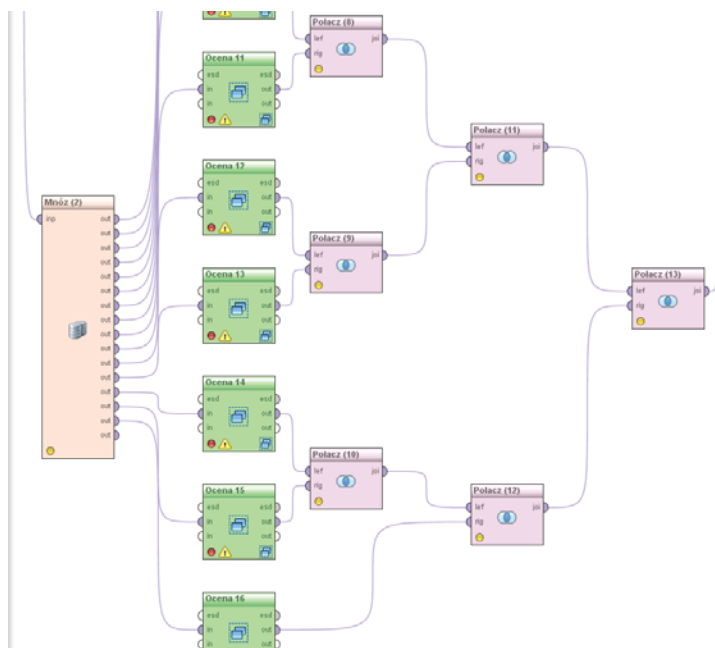
Reguła	"stężenie $CH_4$ "	"przyrost stężenia $CH_4$ "	"stopień zagrożenia"
1	"w normie"	"zerowy"	"normalne"
2	"w normie"	"rosnący"	"normalne"
3	"w normie"	"szybko rosnący"	"ostrzeżenie"
4	"dopuszczalne"	"zerowy"	"ostrzeżenie"
5	"dopuszczalne"	"rosnący"	"ostrzeżenie"
6	"dopuszczalne"	"szybko rosnący"	"zagrożenie"
7	"graniczne"	-	"zagrożenie"
8	"przekroczone"	-	"zagrożenie"

węgla kamiennego. Pod uwagę zostały wzięte te metody, które aktualnie mogą być stosowane na skalę przemysłową i wymagają aparatury, którą kopalnie posiadają, lub w którą mogą być wyposażone. W skład metody kompleksowej wchodzi następujące metody szczególne:

- metoda rozeznania górniczego, umożliwiająca oszacowanie potencjalnego niebezpieczeństwa wystąpienia tąpnięcia ze względu na właściwości pokładu i górotworu, warunki zalegania złoża, zaszczości eksploatacyjne itp. (tąpnięcia pokładowe i stropowe);
- metoda sejsmologiczna, umożliwiająca określanie stanu zagrożenia tąpnięciami stropowymi (monitorowanie stanu zagrożenia) na bazie rejestracji wstrząsów górotworu;

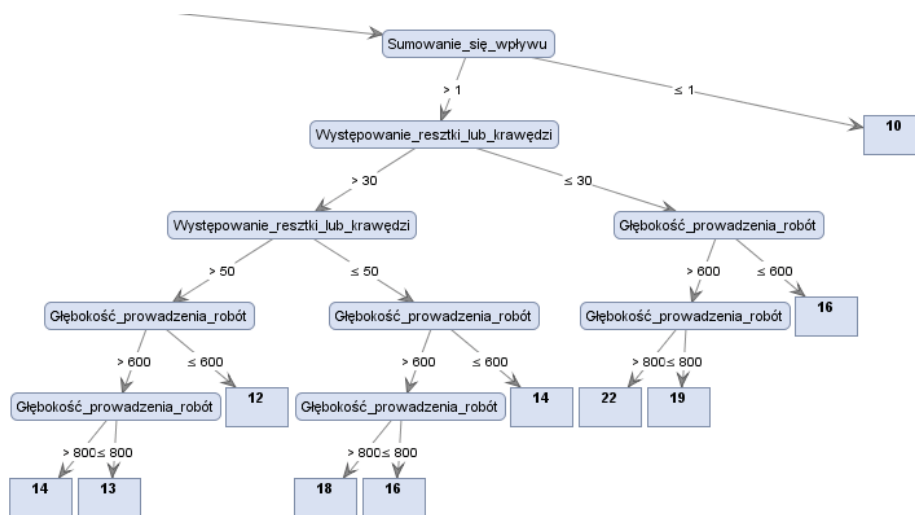
- metoda sejsmoakustyczna, umożliwiająca określanie podwyższonego stanu naprężenia w pokładzie (monitorowanie stanu zagrożenia tąpnięciami pokładowymi);
- metoda energetycznej prognozy zagrożenia sejsmicznego;
- metoda wierceń sondażowych (wierceń małośrednicowych), umożliwiająca określanie podwyższonego stanu naprężenia w pokładzie (monitorowanie stanu zagrożenia tąpnięciami pokładowymi);
- w uzasadnionych przypadkach – inne metody oceny rzeczywistego stanu zagrożenia tąpnięciami.

Przykładowo, czynniki kształtujące stan zagrożenia tąpnięciami i ich wagi uwzględniane w metodzie rozeznania górniczego zostały przedstawione w tabeli 11.2. Czynniki te zostały zaimplementowane w systemie DISESOR. Ilustracja operatorów, które realizują zaimplementowane punkty została przedstawiona na rys. 11.5. Rysunek 11.6 przedstawia natomiast fragment drzewa decyzyjnego, które realizuje metodę rozeznania górniczego, a które jest efektem powstałej implementacji.



Rys. 11.5. Operatory systemu DISESOR (moduł System Ekspertki) realizujące metodę rozeznania górniczego

Metody eksperckie przedstawione powyżej bazowały na wiedzy dziedzinowej. W kolejnej części rozdziału zostaną przedstawione metody bazujące na analizie danych historycznych.



Rys. 11.6. Fragment drzewa decyzyjnego, które realizuje metodę rozeznania górniczego w systemie DISESOR (moduł System Ekspertyki)

Tab. 11.2. Pierwszych 11 (z pośród 15) czynników kształtujących stan zagrożenia tąpnięciami

Lp.	Czynnik wpływający na stan	Definicja czynnika	Waga
1	Głębokość prowadzenia robót	400–600 m	2
		601–900 m	6
		> 900 m	10
2	Występowanie warstw wstrząsogennych w stropie w odległości nad pokładem	20–30 m	5
		30–50 m	3
		< 20 lub > 50 m	2
3	Grubość warstw wstrząsogennych w stropie w odległości do 50 m nad pokładem	10–30 m	5
		$\geq 30$ m	2
4	Występowanie w strefie przodka w odległości do 60 m uskoku o zrzucie $h$	$1,5 \text{ m} < h \leq 3,5 \text{ m}$	1
		$h > 3,5 \text{ m}$	5
5	Sposób prowadzenia przodka w pobliżu uskoku (w odległości do 60 m) o zrzucie $\geq 1,5$ m	zbliżanie się przodkiem do uskoku	5
		przemieszczanie przodka wzdłuż, równoległe do uskoku	3
		oddalanie się przodka uskoku	2
6	Występowanie w stropie lub spągu resztki lub krawędzi w odległości $z$ w pionie przy prowadzeniu robót na głębokości $H \leq 600$ m; przy $600 < H \leq 800$ m podaną wagę mnoży się przez 1,5, a przy $H > 800$ – przez 2	$z \leq 30$ m	6
		$30 < z \leq 50$ m	4
		$50 < z \leq 200$ m	2
7	Sumowanie się wpływu co najmniej dwóch resztek i/lub krawędzi przy najmniejszej odległości w pionie między resztkami i/lub krawędziami 50 m i odległości najbliższej resztki lub krawędzi od pokładu (w przypadku sumowania się wpływów resztek lub krawędzi - nie uwzględnia się lp. 6)	< 30 m	10
		31–50 m	8
		51–100 m	6
8	Roboty górnicze w pokładzie o węgla skłonnym do tępnięć	$30 \text{ MPa} \geq Rc \geq 16 \text{ MPa}$	4
		$16 \text{ MPa} > Rc$ lub $Rc > 30 \text{ MPa}$	1
9	Prognozowana lub stwierdzana maksymalna energia wstrząsu $E$ przy prowadzeniu robót na głębokości $H$	$E < 106 \text{ J}$ przy $H \leq 600 \text{ m}$	4
		$E < 106 \text{ J}$ przy $600 < H \leq 900 \text{ m}$	6
		$E < 106 \text{ J}$ przy $H > 900 \text{ m}$	8
		$E \geq 106 \text{ J}$ przy $H > 400 \text{ m}$	10
10	Wysokość wyrobiska eksploatacyjnego $w$	$3,5 \text{ m} \geq w \geq 2,0 \text{ m}$	4
		$2,0 \text{ m} > w$ lub $w > 3,5 \text{ m}$	2
11	Budowa spodku wyrobiska chodnikowego lub eksploatacyjnego	występowanie w spodku wyrobiska warstwy węgla grubości $\geq 1,0$ m lub podobnej grubości innego pokładu, zalegającego poniżej spągu danego pokładu w odległości $\leq 10$ m	6



### 11.3 Prognozowanie zagrożenia

W niniejszym rozdziale przedstawione zostaną zaimplementowane w systemie DISESOR rozwiązania wykorzystujące metody uczenia maszynowego do prognozowania zagrożeń - metanowego oraz sejsmicznego.

Prognozowanie zagrożenia jest zadaniem trudnym szczególnie ze względu na nierównomierność rozkładu pomiędzy przykładami pozytywnymi („jest zagrożenie”), a przykładami reprezentującymi pozostałe stany (nazwijmy je ogólnie „brak zagrożenia”). Wyzwaniem jest również „problem zimnego startu” [11.24] występujący przy rozpoczynaniu prac przy nowej ścianie oraz fakt, że kolejne ściany mogą różnić się znacząco pod względem liczby występujących sytuacji zagrożenia.

W systemie DISESOR zostały zawarte rozwiązania prognozowania zagrożeń metanowego oraz sejsmicznego, które stanowią adaptację najlepszych rozwiązań zgłoszonych do konkursów analizy danych zorganizowanych przy konferencjach międzynarodowych i ogłoszonych na platformie Knowledge Pit [11.13]. Metody te (oraz wiele innych metod osiągających wysokiej jakości wyniki) charakteryzują się następującymi głównymi etapami analizy:

1. ekstrakcja cech tworzących nową reprezentację danych i będących agregatami analizowanych szeregów czasowych odpornymi na dryf pojęć (ang. concept drift),
2. opracowanie odpowiedniej metody oceny jakości oraz dostrajania stosowanych modeli predykcyjnych,
3. zastosowanie komitetu metod (metod zespołowych, ang. ensemble learning) w celu połączenia prognoz modeli bazowych.

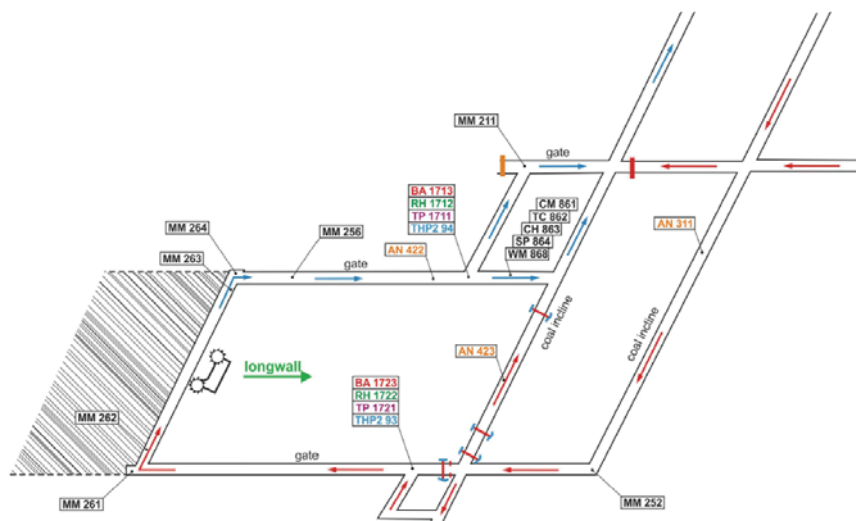
Poniżej zostaną przedstawione założenia dwóch konkursów analizy danych mających na celu wyłonienie adekwatnych metod prognozowania zagrożeń metanowych i sejsmicznych.

#### 11.3.1 Zagrożenie metanowe

Problem prognozowania zagrożenia wynikającego z podwyższonego stężenia metanu w kopalni węgla kamiennego był tematem konkursu analizy danych „IJCRS’15 Data Challenge: Mining Data from Coal Mines” zorganizowanego między 13 kwietnia a 25 czerwca 2015 pod auspicjami konferencji IJCRS’15 (The 2015 International Joint Conference on Rough Sets) [11.9]. Dane konkursowe zawierające odczyty czujników w rejonie przodka zostały dostarczone przez Instytut Technik Innowacyjnych EMAG.

Dla lepszego zrozumienia analizowanych danych została udostępniona mapka pogładowa (rys. 11.7) prezentująca przepływ powietrza w chodnikach oraz położenie czujników. Oprócz pomiarów z czujników gazowych dane zawierały informację na temat prądów elektrycznych pobieranych przez kombajn, ponieważ aktywność kombajnu ma wpływ na ilość uwalnianego metanu.

Każdy przykład danych konkursowych zawierał pomiary dla każdego czujnika rejestrowane co sekundę przez 10 minut (600 pomiarów). Liczba czujników



Rys. 11.7. Schemat rejonu kopalni, dla którego pobrane zostały dane konkursowe

rejestrujących pomiary wynosiła 28. Liczba atrybutów każdego przykładu danych wynosiła więc 16 800. Liczba przykładów to 51 700. Udostępniony zbiór obejmował okres pomiarów pomiędzy 2 marca i 16 czerwca 2014 roku.

Zadaniem konkursowym było prognozowanie czy poziom metanu przekraczający określony próg mógłby wystąpić w ciągu najbliższych 3-6 minut po dokonaniu pomiarów. Przekroczenie stężenia metanu analizowano dla trzech metanometrów: MM263, MM264 and MM256. Jeżeli dany przykład danych odpowiada okresowi czasu  $t_{-599}$  do  $t_0$ , to przykład ten otrzyma etykietę "zagrożenie" wtedy i tylko wtedy, gdy  $(MM(t_{181}), \dots, MM(t_{360})) \geq 1.0$ .

Jest to więc zagadnienie klasyfikacji dwuklasowej ("zagrożenie" i "brak zagrożenia") dla danych o niezrównoważonym rozkładzie klas ("zagrożenie" jest stanem zdecydowanie rzadszym). Ocena wyników dla powyższego zadania została przeprowadzona na podstawie pomiaru pola pod krzywą ROC (ang. Area Under the ROC Curve – AUC) [11.8].

Zaimplementowane w systemie DISESOR rozwiązanie [11.27], które zostało zgłoszone do konkursu i sprawdzone dla opisanych danych składa się z następujących kroków:

1. generowanie zmiennych wywiedzionych,
2. ekstrakcja cech,
3. trening modeli bazowych.

Proces generowania zmiennych wywiedzionych i ekstrakcji cech bazuje na metodzie okna przesuwne (ang. sliding window). Dla każdej pary szeregów czasowych  $x, y$  wyznaczonych przez okno przesuwne na wejściowych atrybutach, które odpowiadają czujnikom metanu lub innym czujnikom środowiskowym

(ciśnienia, wilgotności, temperatury i prędkości powietrza), wyznaczane są nowe szeregi czasowe:

$$z_1 = x - y,$$

$$z_2 = \frac{x-y}{x}.$$

W kolejnym kroku metody dla atrybutów źródłowych oraz wywiedzionych wyznaczone są cechy numeryczne, które podsumowują różne aspekty danych szeregów czasowych. Dla każdego pojedynczego atrybutu wyznaczono 17 cech, w tym m.in.: wartość średnia, minimalna, maksymalna, odchylenie standardowe i inne bardziej złożone charakterystyki. Dla par atrybutów reprezentujących te same typy czujników (np. metanomierz-metanomierz lub barometr-barometr) wyznaczane są również wartości korelacji (w tym autokorelacji) Pearsona, Spearmana, Kendalla.

Dane o tak wyznaczonych cechach wykorzystywane są następnie do wygenerowania modelu bazującego na metodzie lasów losowych [11.7]. Ze względu na zastosowany klasyfikator nie była wymagana selekcja cech.

Omówiona powyżej metoda osiągnęła najlepsze wyniki w ramach przedstawionego konkursu. Uczestnicy konkursu mieli dostęp tylko do wstępnych wyników, które były oparte na podzbiorze końcowego zbioru testowego i odnosiły się do około 20% danych testowych. Ocena końcowa została przeprowadzona po zakończeniu konkursu. Wyniki uzyskane przez najlepszych uczestników konkursu przedstawione są w tabeli 11.3.

Tab. 11.3. Wyniki konkursu prognozowania zagrożenia metanowego

Pozycja	Nazwa	Wstępne wyniki (AUC)	Końcowe wyniki (AUC)
1	Zagorecki	0.9666259	0.95926715
2	March	0.94607103	0.94392893
3	Dymitruta	0.93371596	0.94369948
4	Moomean	0.92862237	0.94280921
5	Trzewior	0.94691336	0.94134706

Model został zaimplementowany jako operator w systemie RapidMiner i może być stosowany w module analitycznym systemu DISESOR. Użytkownik musi wskazać jedynie zbiór zmiennych warunkowych i zmienną decyzyjną (prognozowany metanomierz), oraz horyzont prognozy wynoszący maksymalnie 6 minut, a reszta analizy przebiega automatycznie.

### 11.3.2 Zagrożenie sejsmiczne

Problem prognozowania zagrożeń sejsmicznych w kopalni węgla kamiennego był tematem konkursu analizy danych "AAIA'16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines" zorganizowanego między 5

października 2015 a 27 lutego 2016 pod auspicjami konferencji AAIA'16 (11th International Symposium on Advances in Artificial Intelligence and Applications) wchodzącej w skład konferencji FedCSIS (<https://fedcsis.org/2016/aaia>). Dane konkursowe zostały dostarczone przez Instytut Technik Innowacyjnych EMAG.

Każdy przykład danych konkursowych zawierał godzinne agregaty odczytów sensorów z okresu 24 godzin oraz informację o intensywności prac w danym rejonie, a także ostatnie oceny warunków bezpieczeństwa wyznaczone przez ekspertów górniczych. Ponadto, dla każdego rejonu wydobywania podane zostały meta-dane w postaci identyfikatora kopalni, identyfikatora rejonu oraz głębokości pokładu. Całkowita liczba atrybutów obejmowała 541 atrybutów głównych oraz 6 cech dodatkowych związanych z danym rejonem wydobywania. Całkowita liczba obserwacji tworzących zbiór danych wynosiła 133150. Udostępniony zbiór obejmował sumarycznie pięcioletni okres pomiarów (lata 2010 do 2015).

Zadanie konkursowe polegało na przygotowaniu modelu, który byłby w stanie dokładnie przewidzieć okresy zwiększonej aktywności sejsmicznej. Atrybut decyzyjny przygotowanych danych konkursowych wskazywał przypadki, w których całkowita energia sejsmiczna zarejestrowana w ciągu ośmiu godzin następujących po danym (godzinnym) przedziale agregacji przekraczała poziom ostrzegawczy  $5 * 10^4$  J.

Jest to więc zagadnienie klasyfikacji dwuklasowej ("zagrożenie" i "brak zagrożenia") dla danych o silnie niezrównoważonym rozkładzie klas ("zagrożenie" jest stanem zdecydowanie rzadszym). Horyzont dokonywanej prognozy wynosi 1 godz. Ocena wyników dla powyższego zadania została przeprowadzona na podstawie pomiaru pola pod krzywą ROC (ang. Area Under the ROC Curve – AUC) [11.8].

Zaimplementowane w systemie DISESOR rozwiązanie, które zostało sprawdzone dla opisanych danych, składa się z następujących kroków:

1. generowanie cech,
2. próbkowanie,
3. redukcja cech,
4. filtrowanie podzbiorów cech,
5. trening modeli bazowych,
6. wyznaczenie komitetu modeli bazowych.

Proces ekstrakcji cech bazuje na metodzie okna przesuwanego (ang. sliding window) [11.26] zastosowanego na wielowymiarowym strumieniu danych. Dla wartości atrybutów objętych przez okno przesuwne zastosowanych zostało szereg funkcji agregujących. Dla każdego pojedynczego atrybutu wyznaczono 11 cech, w tym m.in.: wartość średnia, minimalna, maksymalna, odchylenie standardowe i inne bardziej złożone charakterystyki.

Ponadto, wyznaczone cechy dla każdego atrybutu zostały uzupełnione o wartości korelacji Kendall'a dla każdej pary atrybutów. Dodatkowo wyznaczane zostały cechy atrybutów opisujące zmiany wartości pomiędzy następującymi po sobie oknami. Cechy te obejmują m.in. różnicę pomiędzy wartościami maksymalnymi, minimalnymi i średnimi w kolejnych oknach.

Kolejny krok zastosowanej metody wynika z silnej nierówności liczby przykładów oznaczonych jako zagrożenie. Wartość średnia liczby takich przykładów to zaledwie 2,3% wszystkich obserwacji (poszczególne wartości znacząco różnią się w zależności od eksploatowanego wyrobiska). Konieczne więc było zrównoważenie liczby przykładów reprezentujących różne klasy. W tym celu stosowane jest próbkowanie, w wyniku którego liczba przykładów oznaczonych jako zagrożenie stanowić będzie co najmniej 10% liczebności wylosowanego zbioru.

Ze względu na dużą liczbę generowanych cech kolejnym krokiem zastosowanej metody jest selekcja cech [11.11]. Zastosowana metoda filtrowania wywodzi się z teorii zbiorów przybliżonych [11.20] i wykorzystuje pakiet RapidRoughSets [11.14] do wyznaczenia przybliżonych reduktów decyzyjnych [11.12] z pomocą heurystyki DAAR [11.10]. Pakiet RapidRoughSets został opracowany w ramach projektu DISESOR, jego jądrem jest biblioteka RoughSet [11.21]. Przybliżone redukty decyzyjne są względnie małe, dlatego zastosowano łączenie kilku uzyskując pojedynczy podzbiór atrybutów. W efekcie uzyskiwany jest zbiór znacząco różnych podzbiorów atrybutów.

Ostatnie kroki metody polegają na uczeniu modeli predykcyjnych w oparciu o wyznaczone wcześniej różne podzbiory atrybutów oraz na utworzeniu komitetu takich modeli w celu podjęcia ostatecznej decyzji o oznaczeniu nadchodzącego okresu czasu etykietą "zagrożenie". Zastosowane modele predykcyjne to drzewa regresyjne, SVM oraz uogólniony model liniowy.

Omówiona powyżej metoda osiągnęła najlepsze wyniki w ramach przedstawionego konkursu, jednakże nie została ona oficjalnie sklasyfikowana ze względu na to, że została zaproponowana przez jego organizatorów. Najwyżej sklasyfikowana metoda (*tadeusz*), przygotowana przez część zespołu sklasyfikowanego na drugiej pozycji, wykorzystuje starannie wybrany podzbiór danych, dla którego uczono modele predykcyjne wchodzące w skład komitetu. Wyniki uzyskane przez najlepszych uczestników konkursu przedstawione są w tabeli 11.4. Omówiona powyżej metoda przedstawiona jest w niej pod nazwą *snm*.

Tab. 11.4. Wyniki konkursu prognozowania zagrożenia sejsmicznego

Pozycja	Nazwa	Wynik (AUC)
-	snm	0.9396
1	tadeusz	0.9393
2	deepsense.io	0.9384
3	yata	0.9342
4	podludek	0.9336
5	jellyfish	0.9335
...	...	...
18	metoda ekspercka	0.9196

Interesujące, że na 18 pozycji rankingu konkursowego została sklasyfikowana *metoda ekspercka* omówiona w punkcie 11.2.2. Metoda ta bazuje na dwóch ocenach ekspertów górniczych, które zostały podane jako atrybuty danych konkursowych. Pokazuje to, jak trudnym zadaniem dla metod uczenia maszynowego jest prognozowanie zagrożenia sejsmicznego. Jednocześnie jednak widać, że metody bazujące na analizie danych historycznych umożliwiają uzyskanie bardziej dokładnych prognoz niż metody bazujące wyłącznie na wiedzy eksperckiej.

## Literatura

- [11.1] A. Barański, J. Drzewiecki, J. Kabiesz, W. Konopko, J. Kornowski, A. Krzyżowski, G. Mutke. Zasady stosowania metody kompleksowej i metod szczegółowych oceny stanu zagrożenia tąpnięciami w kopalniach węgla kamiennego. *Główny Instytut Górnictwa: Seria Instrukcje*, (20), 2007.
- [11.2] A. Cianciara, B. Cianciara. The meaning of seismoacoustic emission for estimation of time of mining tremors occurrence. *Archives of Mining Sciences*, 51(4):563–575, 2006.
- [11.3] D. W. Dixon. *A statistical analysis of monitored data for methane prediction*. Praca doktorska, University of Nottingham, 1992.
- [11.4] W. Gale, K. Heasley, A. Iannacchione, P. Swanson, P. Hatherly, A. King, i in. Rock damage characterisation from microseismic monitoring. *DC Rocks 2001, The 38th US Symposium on Rock Mechanics (USRMS)*. American Rock Mechanics Association, 2001.
- [11.5] S. J. Gibowicz, S. Lasocki. Seismicity induced by mining: Ten years later. *Advances in geophysics*, 44:39–181, 2001.
- [11.6] T. Grychowski. Hazard assessment based on fuzzy logic. *Archives of Mining Sciences*, 53(4):595–602, 2008.
- [11.7] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
- [11.8] J. A. Hanley, B. J. McNeil. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148(3):839–843, 1983.
- [11.9] A. Janusz, M. Sikora, L. Wróbel, S. Stawicki, M. Grzegorowski, P. Wojtas, D. Slezak. Mining data from coal mines: Ijcrs’15 data challenge. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing - 15th International Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceedings*, s. 429–438, 2015.
- [11.10] A. Janusz, D. Ślęzak. Random probes in computation and assessment of approximate reducts. *International Conference on Rough Sets and Intelligent Systems Paradigms*, s. 53–64. Springer, 2014.
- [11.11] A. Janusz, D. Ślęzak. Rough set methods for attribute clustering and selection. *Applied Artificial Intelligence*, 28(3):220–242, 2014.

- [11.12] A. Janusz, D. Ślęzak. Computation of approximate reducts with dynamically adjusted approximation threshold. *International Symposium on Methodologies for Intelligent Systems*, s. 19–28. Springer, 2015.
- [11.13] A. Janusz, D. Slezak, S. Stawicki, M. Rosiak. Knowledge pit - A data challenge platform. *Proceedings of the 24th International Workshop on Concurrency, Specification and Programming, Rzeszow, Poland, September 28-30, 2015.*, s. 191–195, 2015.
- [11.14] A. Janusz, S. Stawicki, M. Szczuka, D. Ślęzak. Rough set tools for practical data exploration. *International Conference on Rough Sets and Knowledge Technology*, s. 77–86. Springer, 2015.
- [11.15] J. Kornowski. Linear prediction of aggregated seismic and seismoacoustic energy emitted from a mining longwall. *ACTA MONTANA*, 129:5–14, 2003.
- [11.16] E. Kuntz. Gas emission problems in starting the workings and controlling measures. *Kongres*, 2008.
- [11.17] S. Lasocki. Probabilistic analysis of seismic hazard posed by mining induced events. *Proc. Sixth Int. Symp. on Rockburst and Seismicity in Mines "Controlling Seismic Risk*, s. 9–11, 2005.
- [11.18] J. Łęski. *Systemy neuronowo-rozmyte*. Wydawnictwa Naukowo-Techniczne, 2008.
- [11.19] S. Nakayama, K. Uchino, M. Inoue. Simulation of methane gas distribution at a heading face-analysis of environmental conditions at a heading face with auxiliary ventilation (3rd report). *JOURNAL-MINING AND MATERIALS PROCESSING INSTITUTE OF JAPAN*, 114:229–236, 1998.
- [11.20] Z. Pawlak. Rough sets. *International Journal of Computer & Information Sciences*, 11(5):341–356, 1982.
- [11.21] L. S. Riza, A. Janusz, C. Bergmeir, C. Cornelis, F. Herrera, D. Śle, J. M. Benítez, i in. Implementing algorithms of rough set theory and fuzzy rough set theory in the r package "rroughsets". *Information Sciences*, 287:68–89, 2014.
- [11.22] M. Sikora, Z. Krzystanek, B. Bojko, K. Spiechowicz. Hybrid adaptative system of gas concentration prediction in hard-coal mines. *Systems Engineering, 2008. ICSENG'08. 19th International Conference on*, s. 159–164. IEEE, 2008.
- [11.23] M. Sikora, B. Sikora. Application of machine learning for prediction a methane concentration in a coal-mine. *Archives of Mining Sciences*, 51(4):475–492, 2006.
- [11.24] L. H. Son. Dealing with the new user cold-start problem in recommender systems: A comparative review. *Information Systems*, 58:87–104, 2016.
- [11.25] K. Z. Ushakov. Gas dynamics of shafts. *Nauka*, 1984.
- [11.26] A. A. Wieczorkowska, J. Wróblewski, P. Synak, D. Ślęzak, i in. Application of temporal descriptors to musical instrument sound recognition. *Journal of Intelligent Information Systems*, 21(1):71–93, 2003.
- [11.27] A. Zagorecki. Prediction of methane outbreaks in coal mines from multivariate time series using random forest. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, s. 494–500. Springer, 2015.