



ZASTOSOWANIA DATA MINING W SYSTEMIE MONITOROWANIA PRACY KOMBajnÓW GÓRNICZYCH

*Mariusz Gibiec, Akademia Górniczo-Hutnicza im. St. Staszica w Krakowie,
Katedra Robotyki i Mechatroniki*

Kombajny górnicze, podobnie jak wiele innych złożonych maszyn i urządzeń, wyposażane są w różnego rodzaju czujniki i systemy pomiarowe już na etapie ich wytwarzania. Urządzenia pomiarowe stanowią najczęściej elementy systemu sterowania lub systemu monitorowania pracy maszyn. Ich zadaniem jest zapewnienie bezpiecznej eksploatacji. W przypadku złożonych instalacji objętych długoterminową gwarancją celem producenta jest także monitorowanie poprawności eksploatacji pod kątem zgodności z warunkami gwarancji. Takie informacje gromadzone są w dużych bazach danych i poddawane są analizie tylko w przypadku zaistnienia awarii. Jednak ich zawartość może także służyć do określania stanu technicznego urządzenia oraz przyczyn jego zmian. Ponieważ w trakcie eksploatacji gromadzone są wielkie ilości danych, ich analiza wymaga zastosowania metod data mining. Wyniki tych analiz mogą zostać wykorzystane do budowy systemu diagnostyki bez dodatkowych inwestycji sprzętowych. W tym celu można zaproponować pewien proces data mining, którego końcowym efektem jest budowa skutecznego klasyfikatora stanu technicznego maszyny popartego czytelnym mechanizmem wnioskowania. Umożliwia on szybkie określanie przyczyn zmian stanu technicznego oraz symptomów pojawiających się niesprawności.

Analizowane dane

Przedmiotem analizy są dane zbierane podczas pracy kombajnu górniczego. Jest to płytkozabiorowy dwuramieniowy kombajn ścianowy z elektrycznym napędem posuwu do wybierania pokładów węgla. Przystosowany jest do pracy w zmiennych warunkach geologicznych. Stosowany jest do dwukierunkowego, bezwnękowego urabiania i ładowania węgla w ścianowych systemach eksploatacji pokładów nachylonych do 35° w przypadku pokładów podłużnych, a w pokładach poprzecznych do 20° (po wzniesieniu) lub 15° (po upadzie). Za posuw odpowiedzialne są silniki elektryczne zasilane napięciem z falownika w zakresie częstotliwości od 0 do 100Hz. W obszarze do 50Hz regulacja następuje przy stałym momencie, po przekroczeniu progu 50Hz zachowana jest stała moc. Dzięki zastosowaniu przemiennika częstotliwości regulacja prędkości posuwu kombajnu odbywa się automatycznie i bezstopniowo w zależności od obciążenia silników elektrycznych napędów posuwu oraz ramion urabiających. Kombajn ten współpracuje



z systemem sterowania i diagnozowania (w czasie rzeczywistym) pracy kombajnów węglowych o dużej mocy, wyposażonych w układy napędowe składające się z maksymalnie 6 silników elektrycznych. Wszystkie funkcje sterownicze i diagnostyczne są realizowane dzięki informacjom dostarczonym z czujników służących do pomiaru:

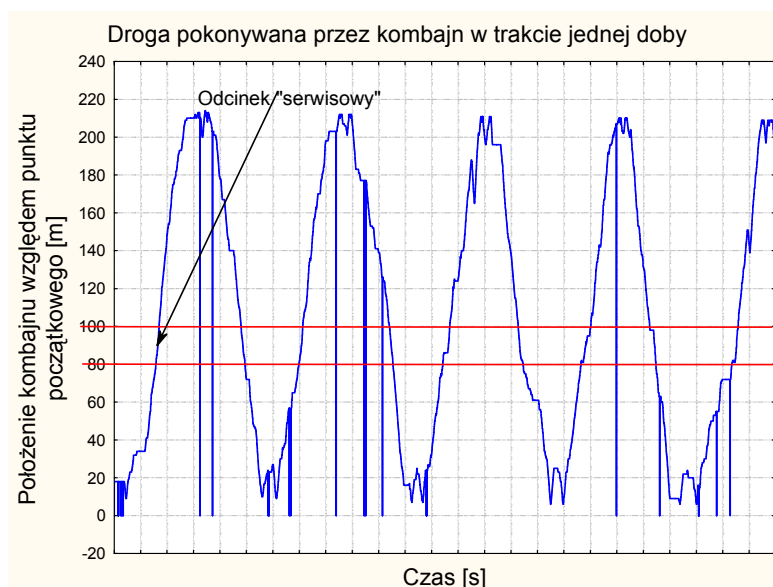
- ♦ prądów obciążenia silników elektrycznych,
- ♦ temperatury uzwojeń i łożysk silników oraz mechanizmów,
- ♦ ciśnienia w obwodach hydraulicznych ciągnika i hamulców,
- ♦ ciśnienia wody chłodzącej,
- ♦ prędkości posuwu kombajnu i położenia kombajnu w ścianie, przez pomiar przebytej drogi.

Diagnostyka oparta na opisywanym systemie może przebiegać na dwa sposoby. Sposób pierwszy polega na bieżącym przetwarzaniu danych pomiarowych i generowaniu komunikatów o stanie kombajnu. Jednak możliwa jest współpraca z układem pomiarowym, który zapisuje parametry związane z pracą kombajnu na przestrzeni 24 godzin z rozdzielczością 1 sekundy. Dane gromadzone są w arkuszach o rozmiarach 127 na „n”, przy czym 127 stanowi liczbę zmiennych uzyskanych z różnych czujników, a druga wartość „n” opisuje liczbę przypadków i zmienia się w zależności od długości pracy kombajnu w danym dniu. Przykład zawartości omówionego pliku danych pokazano na rys. 1.

	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	PradM4	PradM5	PradM6	TempM1	TempM2	TempM3	TempM4	TempM5	TempM6	CislWysokie	CislNiskie	PredkPosuwu	Droga	TempRar
516	37	8	0	67	77	175	69	18	69	0	41	6	99	
517	37	10	0	67	77	175	69	18	69	124	41	6	99	
518	42	7	0	67	77	175	69	18	69	0	41	6	100	
519	51	7	0	50	77	HH5	111	47	114	0	41	6	100	
520	51	7	0	50	77	HH5	111	47	114	0	41	6	100	
521	46	7	0	50	77	HH5	111	47	114	0	41	6	100	
522	51	7	0	HH5	171	25	146	10	94	44	41	6	100	
523	49	7	0	HH5	171	25	146	10	94	0	41	6	100	
524	45	7	0	HH5	171	25	146	10	94	0	41	6	100	
525	53	6	0	HH5	177	65	97	78	168	138	41	6	100	
526	34	6	0	HH5	177	65	97	78	168	0	41	6	100	
527	35	7	0	HH5	177	65	97	78	168	0	41	6	101	
528	45	7	0	51	175	125	125	63	FFF	0	41	6	101	
529	40	7	0	51	175	125	125	63	FFF	0	41	6	101	
530	47	9	0	51	175	125	125	63	FFF	160	41	6	101	
531	55	7	0	31	96	HH5	FFF	20	75	86	41	6	101	
532	66	6	0	31	96	HH5	FFF	20	75	0	41	6	101	
533	48	7	0	31	96	HH5	FFF	20	75	0	41	6	101	
534	50	6	0	44	142	91	84	HH5	142	0	41	6	101	
535	48	10	0	44	142	91	84	HH5	142	52	41	6	101	
536	56	6	0	44	142	91	84	HH5	142	0	41	6	101	
537	51	8	0	41	77	HH5	FFF	1	88	98	41	6	102	
538	54	6	0	41	77	HH5	FFF	1	88	0	41	6	102	
539	45	6	0	41	77	HH5	FFF	1	88	0	41	6	102	
540	41	7	0	47	76	32	128	28	82	0	41	6	102	
541	47	7	0	47	76	32	128	28	82	0	41	6	102	
542	49	6	0	47	76	32	128	28	82	0	41	6	102	
543	51	7	0	40	66	40	128	23	126	0	41	7	102	

Rys. 1. Fragment pliku danych.

Wykorzystywane dane zostały oczyszczone za pomocą narzędzi programu *STATISTICA* i poddane wstępnej analizie. Ponieważ głównym zadaniem systemu diagnostycznego jest klasyfikacja stanu technicznego dokonywana na podstawie mierzonych danych, istotnym jest określenie lub utworzenie zmiennej identyfikującej ten stan. W dostępnym zbiorze danych zmienna o takim charakterze nie występowała. Informacje o zmianach stanu maszyny zawarte były tylko w protokołach serwisowych. Niestety nie zawierają one dokładnego określenia momentu wystąpienia zmiany, zatem mogą być wykorzystane tylko do weryfikacji poprawności działania budowanych klasyfikatorów stanu. Z tego powodu należy dokonać analizy struktury danych, która powinna odpowiadać strukturze zidentyfikowanych stanów maszyny. Na podstawie tej struktury utworzona zostanie zmienna klasyfikująca (zależna) stan. Przeprowadzone analizy różnymi metodami grupowania danych nie ujawniły czytelnej struktury danych. Dokonując analizy kolejnych przypadków, rejestrowanych dla tego samego stanu maszyny, stwierdzono duże rozbieżności między nimi oraz brak okresowości. Może to być spowodowane tym, że kombajn pracuje w środowisku o losowych parametrach, o dużym wpływie na charakter przebiegów rejestrowanych przez czujniki. Dlatego zaproponowano wybór pewnego odcinka pracy, gdzie warunki te są porównywalne. Przeprowadzona analiza danych pozwoliła na wybór odcinka „serwisowego”. Składa się on z kilku sekcji pokonywanych w każdym przejeździe, z taką samą stałą prędkością i bez występowania przeciążeń. Jest on określany na podstawie zmiennej opisującej przebieg zmian położenia kombajnu „w ścianie”. Przykładowy jej przebieg pokazano na rysunku 2.

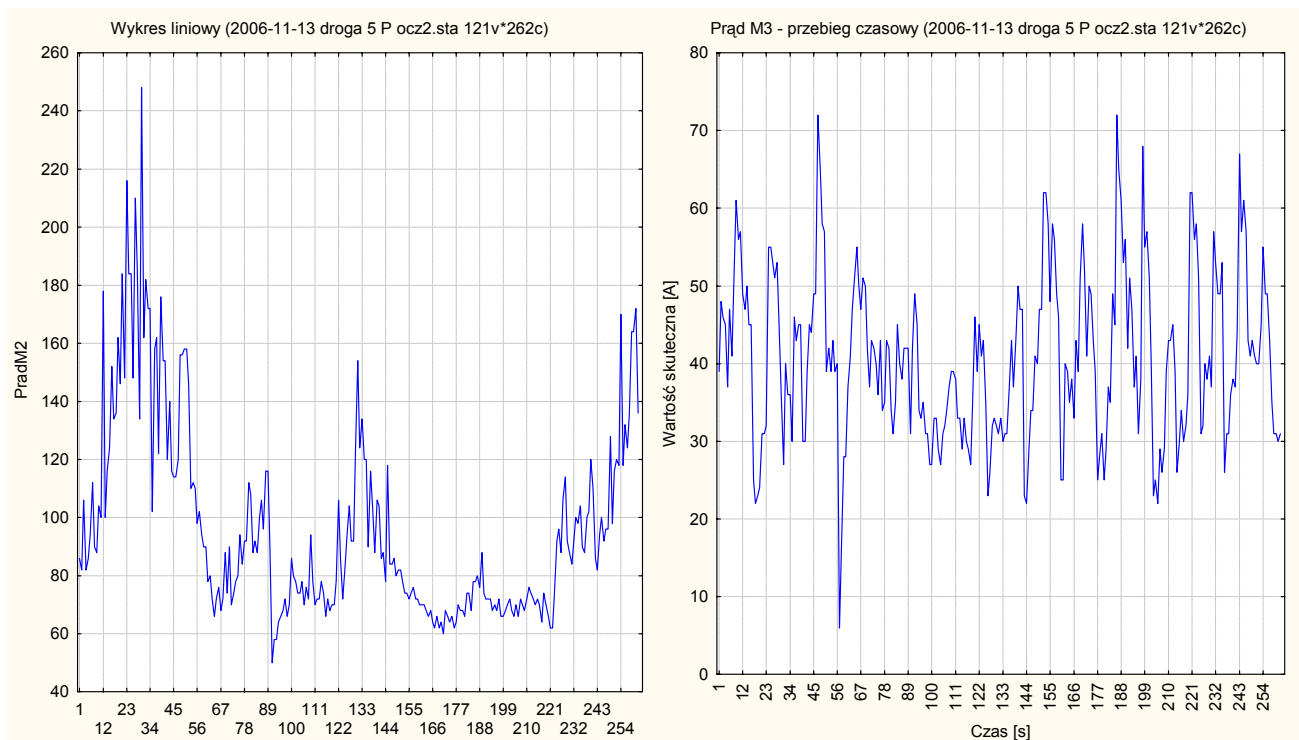


Rys. 2. Położenie kombajnu w funkcji czasu z wyszczególnionym odcinkiem serwisowym.

Duża ilość zmiennych, jak również licznie występujące błędy, były powodem kolejnego etapu filtracji danych. Na podstawie wiedzy uzyskanej od specjalistów, zajmujących się badanym zagadnieniem, zdecydowano się na wyselekcjonowanie następujących zmiennych: „PradM1”, „PradM2”, „PradM3”, „PradM4”, „CzestPrzem”, „PredkSilnPrzem”, „PradSilnPrzem”, „MocSilnPrzem”, „MomentSilnPrzem”, „ACPrzem”, „DCPrzem”. Dodatkowo stwierdzono potrzebę dodania nowej zmiennej „Roznica”, której wartość

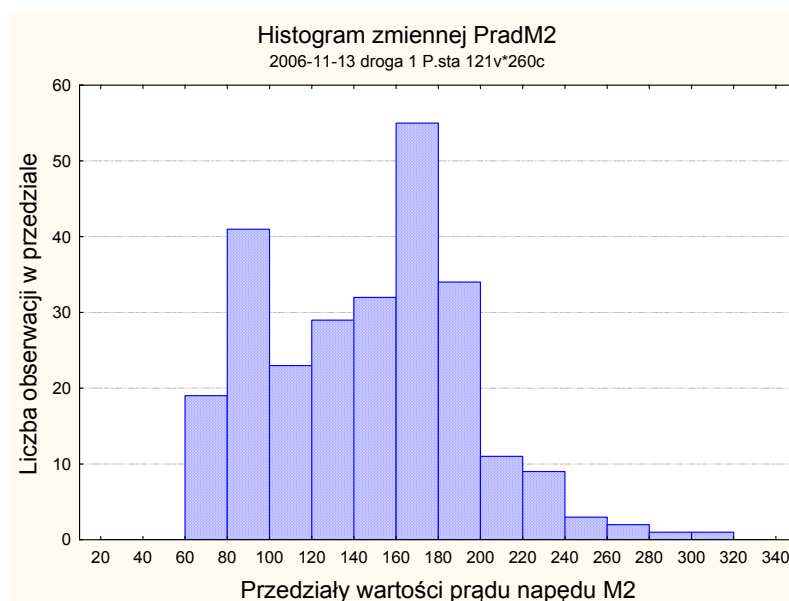


równa jest różnicy pomiędzy zmiennymi „PradM3” i „PradM4”. Na rys. 3 przedstawiono przykładowe wykresy zmiennych „PradM2”, „PradM3”.



Rys. 3. Przebiegi zmiennych „PradM2”, „PradM3”.

Ponieważ uzyskane charakterystyki wykazują dużą zmienność nawet dla kolejnych przejazdów, zaproponowano zastosowanie metody ilościowego badania powstałych odcinków danych w postaci analizy histogramów. Histogramy te znormalizowano z powodu różnej ilości zebranych danych dla tego samego odcinka w różnych przejazdach.

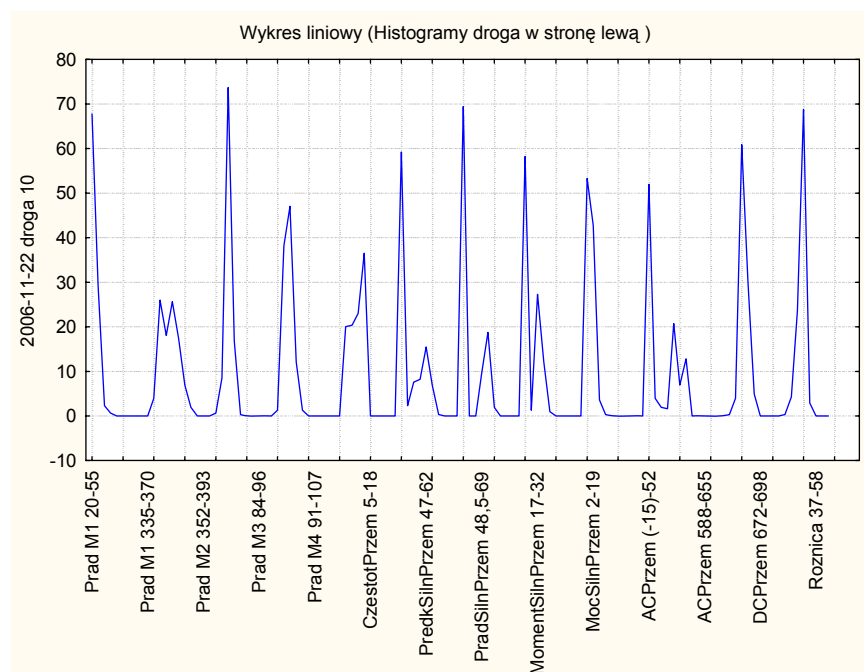


Rys. 4. Histogram zmian prądu silnika napędu.



Na rys. 4 pokazano przykładowy histogram dla zmian prądu silnika napędu.

Histogramy dla wszystkich zmiennych połączono w jeden wektor, stanowiący nową charakterystykę pojedynczego przejazdu. Wykres takiego wektora przypomina wyglądem widmo sygnału (rys. 5). Tego typu dane były wykorzystywane z powodzeniem do klasyfikacji stanu maszyn wirnikowych opisanej w pracy [2]. Ponieważ tak skonstruowane wektory danych dla kolejnych przypadków są podobne, stwierdzono, że tak przekształcone dane będą stanowiły podstawę dalszych analiz.



Rys. 5. Charakterystyka połączonych histogramów.

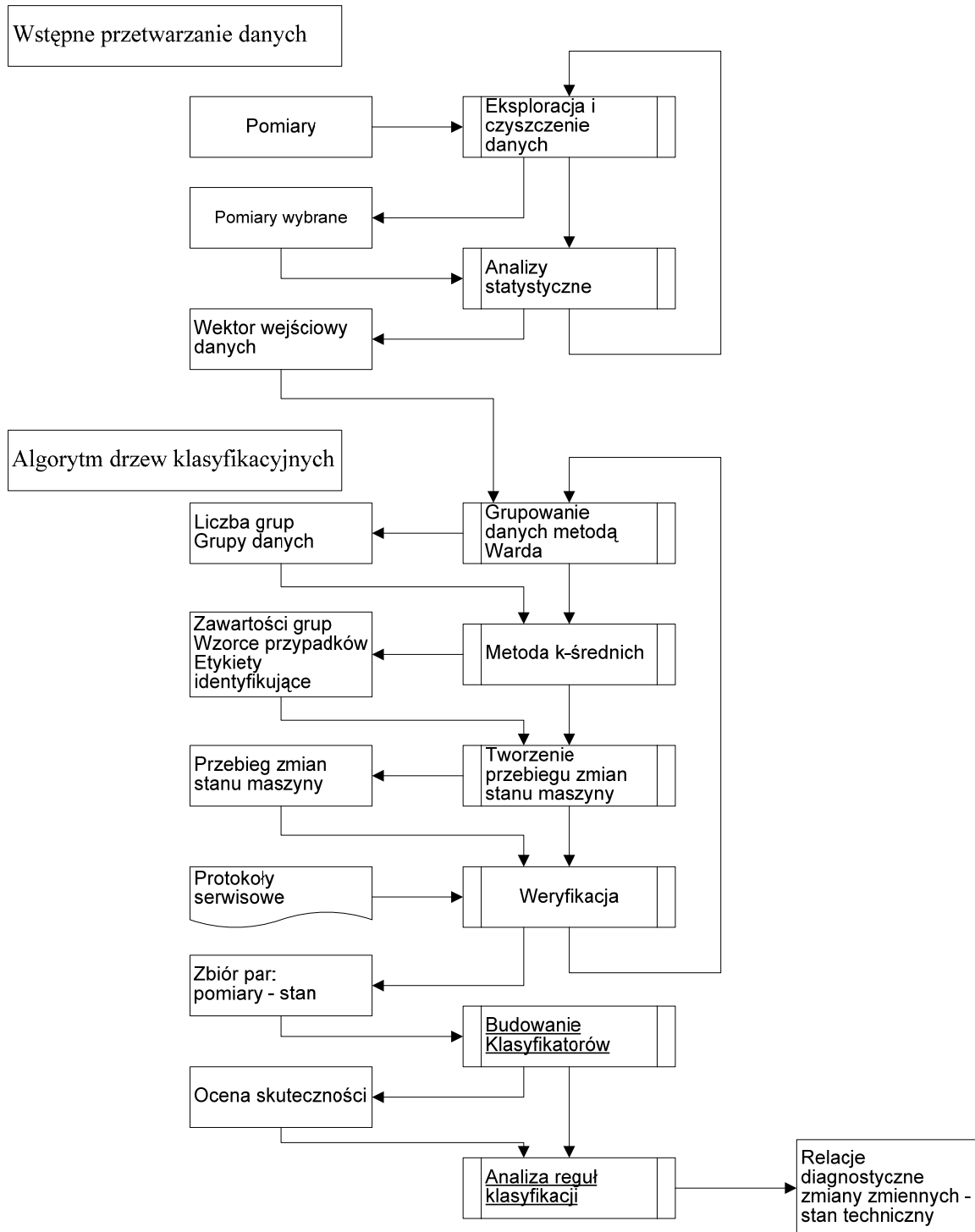
Budowa systemu diagnostycznego

Sposób budowy systemu diagnostycznego, którego głównym elementem jest klasyfikator stanu technicznego maszyny, można opisać jako pewien proces składający się z dwóch głównych części oraz pętli, umożliwiających powrót do wcześniej wykonanych etapów procesu. Pierwsza część dotyczy wstępnego przetwarzania i analizy danych i została opisana wcześniej. W drugiej części następuje budowa klasyfikatorów stanu i wybór najlepszego z nich oraz budowa drugiego elementu systemu diagnostycznego, jakim jest baza wiedzy zawierająca reguły klasyfikacji. W tej części wybrane techniki data mining wykorzystano do odkrywania relacji diagnostycznych na podstawie danych z rejestratora pracy kombajnu górniczego. W pierwszym etapie dokonano grupowania danych w celu określenia, jakie uszkodzenia lub stany pracy kombajnu znajdują swoje odzwierciedlenie w danych. Weryfikacji uzyskanych grup dokonano w oparciu o protokoły serwisowe. Zawierają one opis czynności serwisowych wykonanych bezpośrednio po okresie, w którym zarejestrowano dane. Określone w ten sposób związki grupa danych – stan techniczny stanowią podstawę do budowy modeli klasyfikujących stan techniczny.



Zastosowane metody drzew klasyfikacyjnych pozwalają na interpretację działania mechanizmu klasyfikacji w postaci reguł „jeżeli..., to...” W ten sposób mogą zostać pozyskane relacje diagnostyczne z danych opisujących historię eksploatacji.

Na rys. 6 przedstawiono schemat procesu budowy wymienionych wcześniej elementów systemu diagnostycznego.

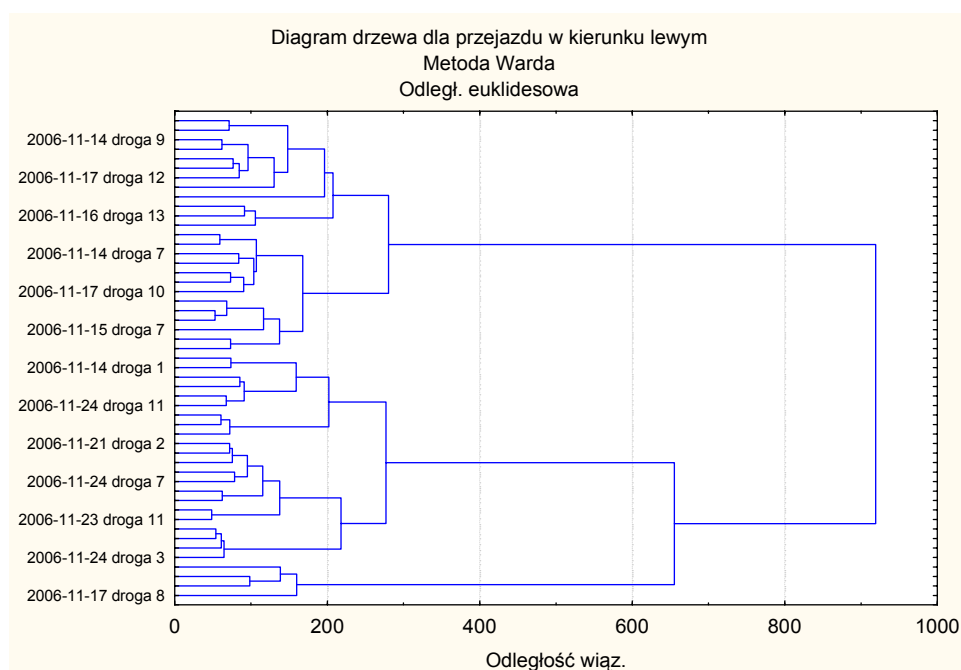


Rys. 6. Schemat budowy systemu diagnostycznego.



Określanie struktury danych

W pierwszym etapie przystąpiono do określenia liczby grup występujących w danych. Liczba grup sugeruje, ile stanów technicznych urządzenia można wyróżnić na podstawie zgromadzonych danych. W analizach założono brak informacji o liczbie grup – zebrane dane pomiarowe nie zawsze są skorelowane z informacjami na temat stanu technicznego. Do grupowania zaproponowano metodę Warda. Jest to metoda z grupy hierarchicznych metod aglomeracyjnych. Tego typu metody pozwalają na określenie tzw. hierarchii drzewkowej elementów analizowanego zbioru. Drzewo połączeń otrzymuje się poprzez krokowe łączenie w podzbiory operacyjnych jednostek taksonomicznych. Na wstępie przyjmuje się, że każdy element zbioru stanowi taką jednostkę. W utworzonej macierzy odległości między jednostkami wyszukuje się najmniejszy element pośród leżących poza przekątną. Jest to odległość aglomeracyjna, minimalna w sensie „lokalnym”. Wskazane przez nią jednostki zostają połączone, tworząc nową jednostkę. Następnie korygowana jest macierz odległości i procedura jest powtarzana. Warunkiem stopu jest uzyskanie jednej jednostki taksonomicznej. Metoda Warda różni się od pozostałych metod sposobem szacowania odległości między jednostkami taksonomicznymi. Wykorzystuje ona analizę wariancji – zmierza do minimalizacji sumy kwadratów odległości dowolnych dwóch skupień, które są tworzone na każdym etapie aglomeracji. Zaletą wybranej metody jest brak konieczności arbitralnego definiowania liczby grup w analizowanych danych oraz o 40% lepsza efektywność wykrywania prawdziwej struktury danych niż w innych metodach [2]. Poniżej zaprezentowano wyniki grupowania dla wektora wejściowego utworzonego z połączonych histogramów wszystkich zmiennych.



Rys. 7. Wyniki grupowania metodą Warda.

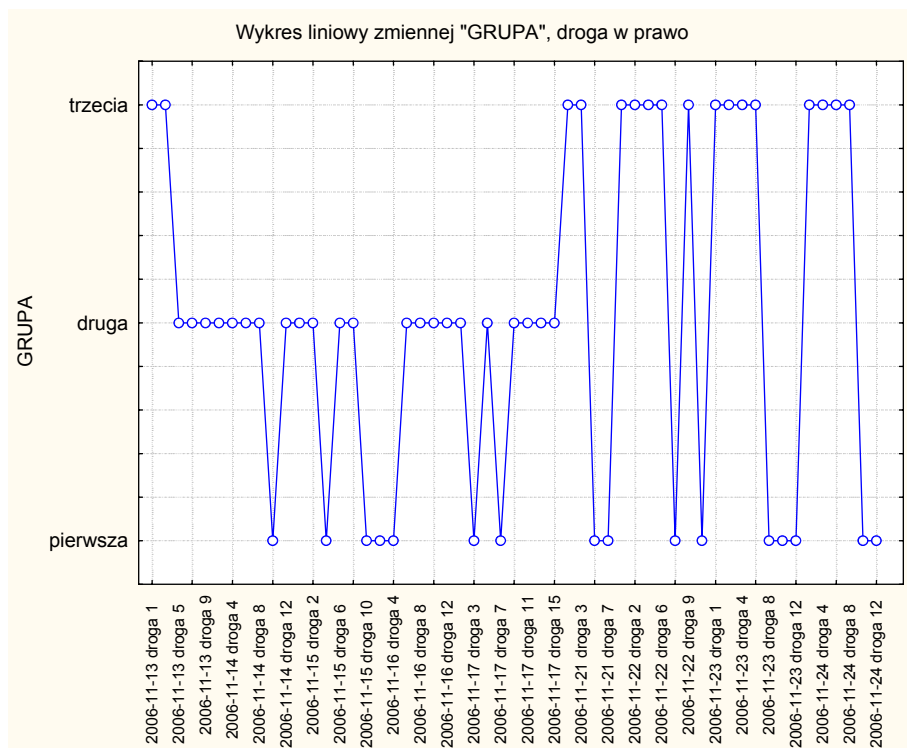


Dokonując przecięcia drzewa Warda na poziomie odległości wiązania równej 400, uzyskujemy trzy grupy danych. Z wykresu można odczytać, które przypadki należą do tej samej grupy.

Na podstawie tego wyniku przeprowadzono grupowanie metodą k-średnich na trzy grupy. Metoda ta różni się od metod aglomeracji założeniem określonej liczby grup w danych. Jej celem jest utworzenie zadanej liczby możliwie odmiennych skupisk danych. W pewnym sensie metoda ta jest odwrotnością analizy wariancji. Działanie rozpoczyna się od losowo wybranych skupisk, a następnie przenosi elementy zbioru między skupiskami, tak by zapewnić minimalizację zmienności wewnątrz skupisk i maksymalizację zmienności pomiędzy nimi.

W analizowanym przykładzie celem zastosowania metody k-średnich było określenie, jakie przypadki tworzą poszczególne grupy. Ich przynależność określa nowa zmienna klasyfikująca „GRUPA”.

Wykorzystując dostępne protokoły serwisowe, przeprowadzono analizę zmian wartości zmiennej klasyfikującej w odniesieniu do przeprowadzonych czynności serwisowych. Na rys. 8 przedstawiono wykresy zmian wartości nowo powstałej zmiennej względem kolejnych przejazdów przez odcinek serwisowy.



Rys. 8. Wartości zmiennej klasyfikującej w kolejnych przejazdach.

Na podstawie otrzymanego opisu czynności serwisowych, przeprowadzonych w rozpatrywanym okresie, odpowiednim grupom przyporządkowano stany maszyny. Jeżeli zmienna grupa przyjmuje wartość *Pierwsza*, to można zakwalifikować stan urządzenia jako poprawną pracę kombajnu, wartość *Druga* powinna odwzorowywać zbiór przypadków,



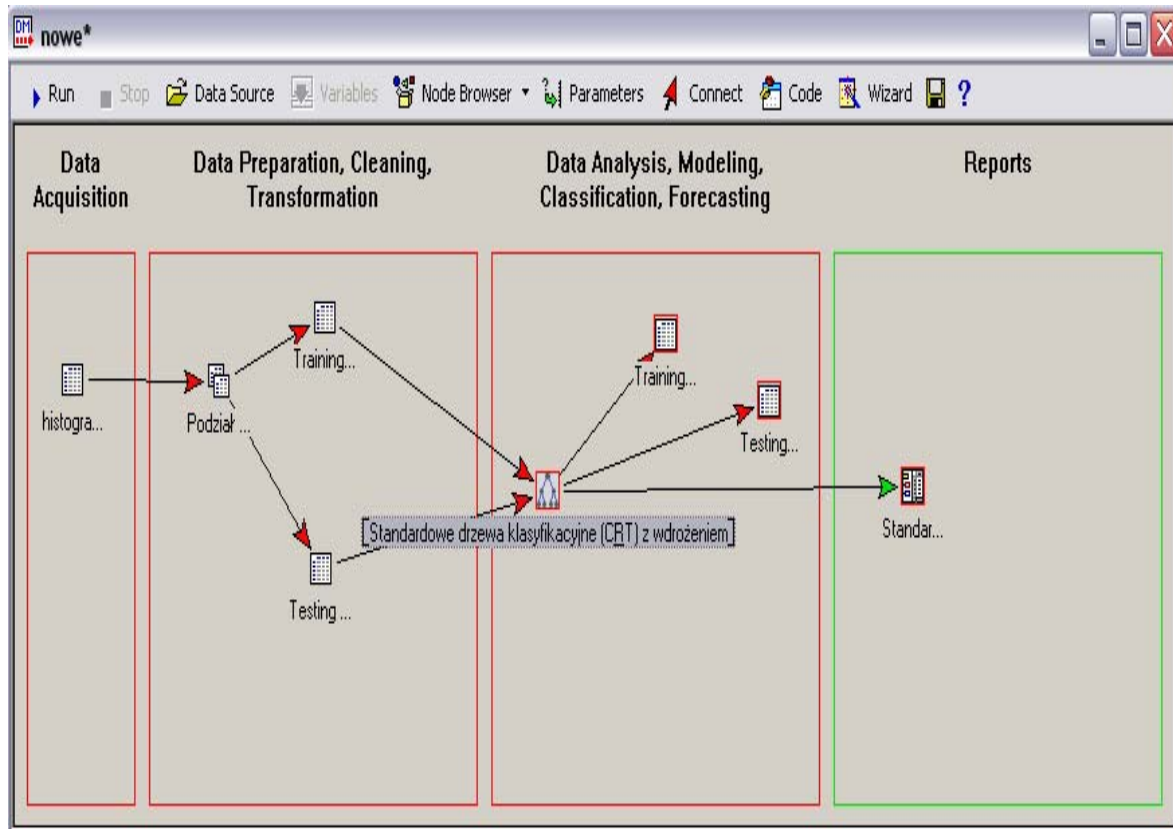
w których występowały uszkodzenia. Opinia ta oparta jest na zapisie wystąpienia w dniu 17.11.2006 r. uszkodzenia koła napędowego lewego napędu, a wartość *Druga* pojawiła się w dniu 16.11.2006 r. i zmieniła się w następnym dniu po wznowieniu pracy kombajnu po naprawie. Według opisu zawartego w części serwisowej wymiany uszkodzonego koła napędowego lewego ciągnika dokonano w dniu 18.11.2006 r. W przypadku wartości *Trzecia* można stwierdzić, że powinna ona również odwzorowywać poprawną pracę urządzenia, lecz jest to inny stan w porównaniu do wartości *Pierwsza*, ponieważ podczas postoju w dniu 20.11.2006 wykonano czynności serwisowe oraz wymieniono uszkodzone części. „Wymieniono przekaźniki czasowe PC, zastąpiono je przekaźnikami RTx-410, przekaźnik K1 zastąpiono przekaźnikiem CI-4. Stwierdzono nieprawidłowości działania zabezpieczeń silników AMP wynikające z fizycznych uszkodzeń przewodów z przekładników pomiarowych”[10]. Jednocześnie można przyjąć, że wymienione wyżej uszkodzenia nie znalazły odzwierciedlenia w analizowanych danych.

Klasyfikacja stanu technicznego

Dysponując utworzonymi w sposób opisany wcześniej przykładami klasyfikacji stanu technicznego, przystąpiono do budowy modeli klasyfikujących. Dla wszystkich metod dokonano podziału zgromadzonych przykładów na próbę uczącą i testową. Ponieważ przykłady do dalszych analiz są wybierane losowo, określono tylko przybliżoną liczbę przykładów testowych na 30% wszystkich danych wejściowych.

Pierwszą z rozważanych metod klasyfikacyjnych jest model oparty na metodzie „Standardowych drzew klasyfikacyjnych C&RT”. Graficznym wynikiem podziału zbioru danych jest drzewo. Powstaje ono wskutek rekurencyjnego podziału zbioru obserwacji A na n rozłącznych podzbiorów $A_1, A_2, A_3, \dots, A_n$. Budowa modelu ma na celu wypracowanie podzbiorów maksymalnie jednorodnych z punktu widzenia wartości zmiennej zależnej. W kolejnych etapach budowy modelu analizowane są wszystkie predyktory i wybierany jest ten, który umożliwia najlepszy podział węzła, co ma prowadzić do powstania najbardziej homogenicznego podzbioru [9].

Każdy model drzewa rozpoczyna się od całego zbioru obserwacji. Zbiór ten ulega podziałowi na dwa (drzewa binarne) lub więcej (drzewa dowolne) podzbiory. Powstałe węzły nazywane są węzłami potomkami (ang. *child nodes*), a wydzielone zostały z tzw. węzła macierzystego (ang. *parent node*). Jeżeli w następnym etapie nie nastąpi kolejny podział węzła potomka, staje się on węzłem końcowym (inaczej zwanym liściem). Jednakże jeśli w drugim etapie węzeł potomek ulega kolejnemu podziałowi, staje się węzłem macierzystym dla danego etapu, a nowo powstałe węzły nazywane są potomkami. Metoda ta umożliwia odkrywanie pewnych reguł występujących pomiędzy zmiennymi w badanym zbiorze danych wejściowych. Opracowany w środowisku *STATISTICA* projekt *Data Mining* pokazano na rys. 9.

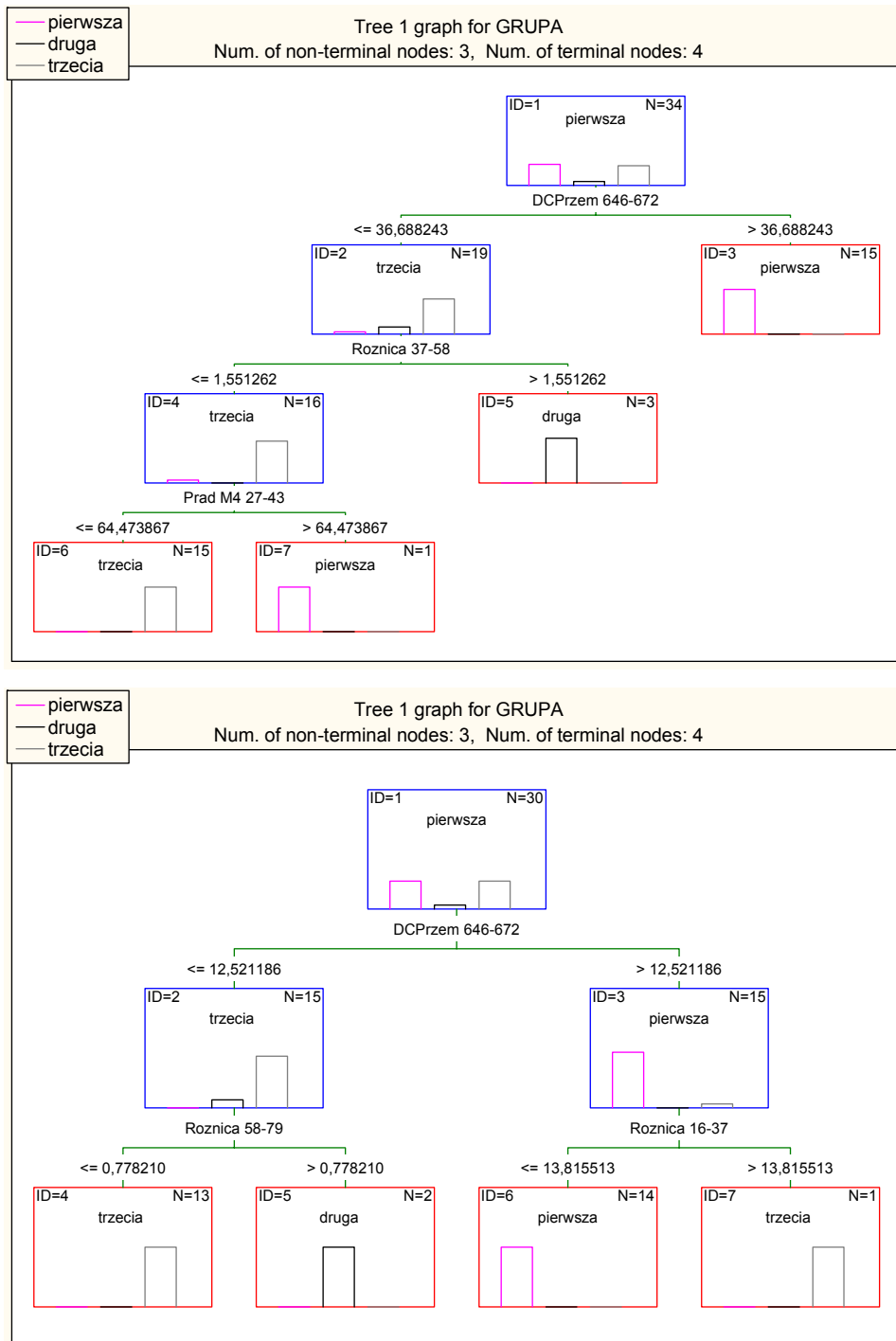


Rys. 9. Model klasyfikacyjny oparty na standardowych drzewach klasyfikacyjnych C&RT.

Wyniki budowy modeli klasyfikujących w postaci drzew klasyfikacyjnych przedstawiono na rysunkach 10-11. Z tego powodu, iż przypadki w zbiorze uczącym wybierane są w sposób losowy, otrzymano kilka drzew. Przy każdorazowym uruchomieniu tworzenia modelu otrzymano nowe drzewo, które w różnym stopniu odwzorowuje relacje występujące w danych. Poniżej przedstawione są tylko te drzewa, dla których błąd dopasowania był najmniejszy.

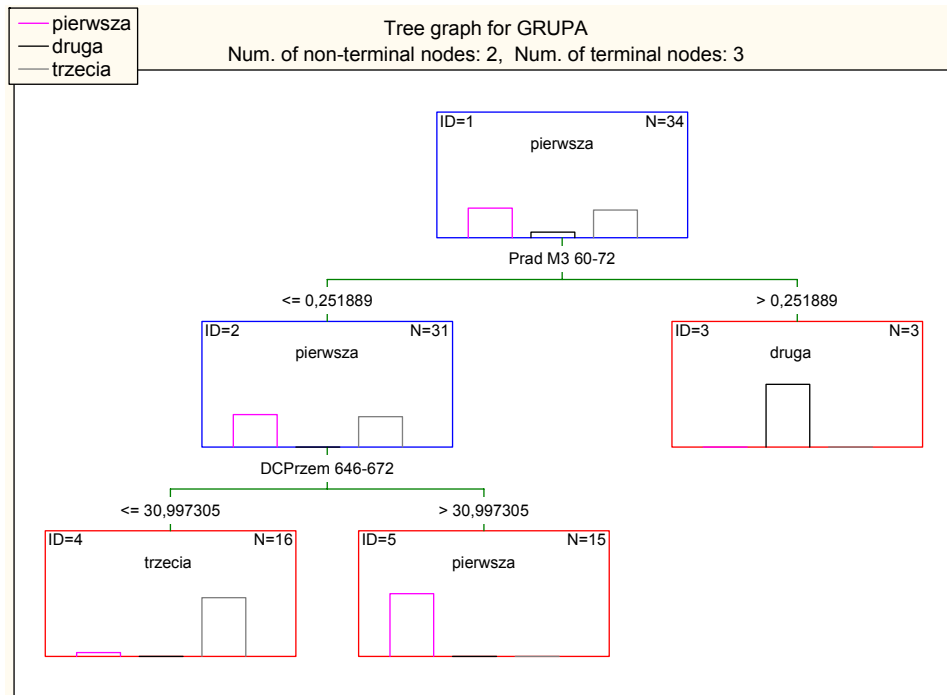
Na przykładzie pierwszego drzewa można określić następujące reguły opisujące zależności pomiędzy opracowanymi danymi i stanem technicznym kombajnu:

1. Węzeł nr 2: „jeśli zmienna DCPrzem594-620 przyjmuje wartość mniejszą bądź równą 2,862957, to stan techniczny należy do pierwszej grupy”.
2. Węzeł nr 4: „jeśli zmienna DCPrzem594-620 przyjmuje wartość większą od 2,862957 i zmienna Roznica37-58 przyjmuje wartości mniejsze bądź równe 1,551262, to stan techniczny należy do trzeciej grupy”.
3. Węzeł nr 5: „jeśli zmienna DCPrzem594-620 przyjmuje wartość większą od 2,862957 i zmienna Roznica37-58 przyjmuje wartości większe od 1,551262, to stan techniczny należy do drugiej grupy”.



Rys. 10-11. Przykładowe wyniki C&RT.

Następny model został zbudowany na podstawie metody standardowej klasyfikacji CHAID. Drzewa tego typu wykorzystują wyniki testu Chi-kwadrat jako kryterium podziału węzła. Umożliwiają także podział na więcej niż dwie kategorie w jednym węźle. Metoda ta wykorzystuje jakościowe zmienne wejściowe, zatem konieczne było utworzenie kategorii, do których zakwalifikowano procentowe wartości opracowanych histogramów. Uzyskane drzewo klasyfikacyjne zaprezentowano na rysunku 12. Realizujące mechanizm podziału reguły diagnostyczne mogą zostać odczytane w sposób opisany powyżej.



Rys. 12. Drzewo klasyfikacyjne typu CHAID.

Ponieważ wszystkie modele klasyfikujące są obarczone pewnym błędem, istotny jest wybór najlepszego z nich. W tym celu bada się jego zdolność do generalizacji, wykorzystując zbiór przykładów testowych. Kryterium oceny w dostępnym oprogramowaniu jest niezgodność procentowa określająca procent błędnych klasyfikacji danego modelu dla tej samej próbki testowej. Wyniki dla testowanych modeli oraz modeli neuronowych, wykonanych dla porównania poprawności klasyfikacji, zestawiono w tabeli 1. Wartości te można wykorzystać do określenia stopnia ufności wobec relacji diagnostycznych odkrytych za pomocą danego modelu.

Tabela 1. Porównanie niedokładności klasyfikacji.

Metoda klasyfikacji	Miara oceny poprawności klasyfikacji		
	Statystyka Chi-kwadrat	Statystyka G-kwadrat	Niezgodność procentowa
Sieć neuronowa	49,5	37,73394	55,0
Wyczerpujący CHAID	1,5	7,23546	15,0
Standardowy CHAID	1,5	7,23546	15,0
Standardowy C&RT	1,0	2,772599	5,0



Podsumowanie

W rozważanym przykładzie producent kombajnu górniczego udostępnił tylko krótki 10-dniowy zapis danych opisujących eksploatację kombajnu górniczego. Sposób rejestracji danych wymusił dokonanie wstępnego przetworzenia danych na potrzeby metod data mining. Wykorzystano dwa typy technik data mining. W pierwszym etapie określono, ile stanów technicznych urządzenia znajduje odzwierciedlenie w danych pomiarowych. Ponieważ nie wszystkie zmiany stanu technicznego muszą implikować zmiany wartości mierzonych wielkości weryfikacji ilości zmian stanu technicznego dokonano w oparciu o zapisy w protokołach serwisowych. Wykorzystując tak utworzone związki dane – stan, zbudowano modele klasyfikujące w postaci drzew decyzyjnych. Ich zastosowanie daje możliwość odkrywania relacji diagnostycznych w zgromadzonych danych w postaci czytelnych i zrozumiałych reguł. Wiarygodność odkrytych reguł określono na podstawie dokładności klasyfikacji zastosowanych modeli. W przypadku rozważanego systemu sterowania i monitorowania pracy kombajnu ścianowego przedstawione metody pozwalają na budowę systemu diagnostyki w postaci klasyfikatorów stanu i bazy wiedzy opisującej reguły klasyfikacji, bez dodatkowych inwestycji sprzętowych.

Literatura

1. Grabiński T., Sokołowski A.: The Effectiveness of Some Signal Identification Procedures, Signal Processing: Theories and Applications, North-Holland Publishing Company, EURASIP, 1980.
2. Gibiec M.: Soft Computing tools for machine diagnosing, Journal of Theoretical and Applied Mechanics. 3, vol. 42: 483–501, 2004.
3. Hand D., Mannila H., Smyth P.: Principles of Data Mining, MIT Press, Cambridge. Tłum pol. Eksploracja Danych, WNT, Warszawa 2005.
4. Kantardzic M.: Data Mining: Concepts, Models, Methods and Algorithms, Wiley-Interscience, Hoboken NJ 2003
5. Larose D.: Data Mining Methods and Models. Wiley-Interscience, Hoboken NJ 2006
6. Wang X. Z., Data Mining and Knowledge Discovery for Process Monitoring and Control, Springer-Verlag London 1999.
7. Sohn H., Worden K., Farrar C. R. (2002), Statistical damage classification under changing environmental and operational conditions, Journal of Intelligent Material Systems and Structures, 13 561-574.
8. Skormin V. A., Popyack L. J., Gorodetski V. I., Araiza M. L., Michel J. D. (1999), Applications of cluster analysis in diagnostics-related problems, in: Proceedings of the 1999 IEEE Aerospace Conference, Vol. 3, Snowmass at Aspen, CO, USA, pp. 161-168.
9. Wang K. (2003), Intelligent condition monitoring and diagnosis systems, a computational intelligence approach. ISSN: 0922-6389, IOS Press.