

# WSPOMAGANIE PROJEKTOWANIA PROCESU TECHNOLOGICZNEGO PRZY UŻYCIU DRZEW KLASYFIKACYJNYCH

Izabela ROJEK

**Streszczenie:** W pracy przedstawiono modele doboru wspomagające projektowanie procesu technologicznego przy użyciu drzew klasyfikacyjnych. Modele te wspomagają dobór obrabiarki, narzędzi i parametrów obróbki. Zostały one opracowane przy użyciu wybranych drzew klasyfikacyjnych, takich jak: ogólne modele drzew klasyfikacyjnych (C&RT), ogólne modele CHAID, wzmacniane drzewa klasyfikacyjne oraz las losowy. Wykonano je przy użyciu oprogramowania StatSoft Statistica Data Miner.

**Słowa kluczowe:** drzewo klasyfikacyjne, obrabiarka, narzędzie, parametry obróbki

## 1. Wprowadzenie

Proces technologiczny w budowie maszyn jest podstawową częścią procesu produkcyjnego związaną bezpośrednio ze zmianą kształtu, wymiarów, jakości powierzchni i własności fizyko-chemicznych przedmiotu obrabianego. Z definicji tej wynika funkcja procesu technologicznego, jaką jest zmiana stanu przedmiotu obrabianego ze stanu początkowego (będącego stanem półfabrykatu lub materiału wyjściowego) na stan końcowy (gotowy wyrób) [1,2].

Tradycyjne podejście w projektowaniu procesu technologicznego polega na analizie rysunku części, metod wykonania oraz identyfikacji i porównywaniu technologii części o podobnych elementach geometrycznych, a następnie opracowaniu procesu technologicznego. Proces technologiczny zostaje najczęściej opracowany na podstawie wiedzy o procesach technologicznych podobnych części. Wtedy charakteryzuje go długi czas projektowania oraz znaczny udział prac rutynowych i czasochłonnych.

Projektowanie procesów technologicznych utraciło tradycyjny charakter wraz z możliwością zastosowania technik informatycznych w technologicznym przygotowaniu produkcji. Wpływ na rozwój metod projektowania procesów technologicznych miały również automatyzacja, zasady typizacji oraz technologii grupowej [1,2].

Na przestrzeni lat zmianie ulegały podejścia do projektowania procesu technologicznego. Powszechne stało się wykorzystanie baz danych gotowych procesów technologicznych (systematyczne przeszukiwanie wcześniejszych rozwiązań, zastosowanie wariantowości rozwiązań technologii grupowej, danych przyszłościowych do modelowania i symulacji zaprojektowanych procesów) oraz zastosowanie systemów CAD, CAPP oraz ich integracji (poprzez wspólne bazy danych, bazy wiedzy) [3,4].

Pierwsze technologie sztucznej inteligencji wykorzystywane w systemach CAPP obejmowały systemy ekspertowe z bazą wiedzy w postaci ram, reguł decyzyjnych, sieci semantycznych. Współczesne technologie występują w postaci maszynowego uczenia, logiki rozmytej, sieci neuronowych, algorytmów genetycznych oraz systemów hybrydowych [5].

Zastosowanie maszynowego uczenia w systemach CAPP pozwala na zawarcie doświadczenia technologa w postaci wiedzy w bazach wiedzy oraz na przeprowadzenie w trakcie projektowania procesu technologicznego wnioskowania zbliżonego do rozumowania człowieka.

Praca jest kontynuacją badań częściowo opisanych w artykule [6]. Obecne badania rozszerzają parametry doboru narzędzi. Przeprowadzono badania dotyczące opracowania modeli doboru obrabiarki i parametrów obróbki oraz zastosowano inne drzewa klasyfikacyjne do budowy modeli.

Praca zawiera kilka rozdziałów. Pierwszy jest wprowadzeniem do wspomagania projektowania procesu technologicznego. Drugi zawiera opis danych i wiedzy technologicznej. Trzeci dotyczy opisu drzew klasyfikacyjnych i metody indukcji tych drzew, które zostały użyte do opracowania modeli doboru obrabiarki, narzędzi oraz parametrów obróbki. Czwarty rozdział omawia eksperymenty i opracowane modele. Zostały zastosowane i porównane wybrane algorytmy drzew klasyfikacyjnych. Modele wykonano przy użyciu oprogramowania StatSoft Statistica Data Miner. Ostatni rozdział to podsumowanie.

## **2. Dane i wiedza technologiczna**

Rozpoznanie struktury organizacyjnej przedsiębiorstwa oraz przebiegu produkcji wyrobów jest pierwszym etapem akwizycji danych i wiedzy na potrzeby inteligentnego systemu CAPP.

Sposób projektowania procesu technologicznego oraz zakres zbieranej informacji zależny jest również od rodzaju produkcji przedsiębiorstwa. Rozpatrywany rodzaj produkcji to produkcja jednostkowa i małoseryjna. Charakterystyczną cechą obu produkcji jest duża liczba wariantów wyrobów produkowanych w niewielkich seriach lub jednostkowo. Ich różnorodność powoduje, że stosowany jest niewielki stopień standaryzacji.

Dane produkcyjne w przedsiębiorstwie koncentrują się wokół trzech podstawowych grup obiektów (zleceń, wyrobów, środków produkcji) [4]. Obiektowi zlecenia przyporządkowane są dane o terminach i ilości produkowanych wyrobów. Wokół obiektu wyroby koncentrują się wszystkie dane, które go opisują, tzn. struktura, geometria, materiał, dane technologiczne, proces technologiczny oraz dane dotyczące cyklu życia wyrobu. Natomiast obiekt środki produkcji zawiera dane, które opisują możliwości produkcyjne przedsiębiorstwa (maszyny, oprzyrządowanie, narzędzia). Dane o zleceniach, wyrobach oraz środkach produkcji wykorzystywane są w trakcie projektowania procesu technologicznego. Technolog korzysta z informacji charakteryzujących się dużą liczbą danych, które ulegają ciągłej modyfikacji. Dane te poszukiwane są według określonych kryteriów. Ponadto informacje występują pod różnorodną postacią (tekst, rysunki), w różnych miejscach w przedsiębiorstwie. Dlatego ważnym staje się szybki dostęp do danych prawdziwych i aktualnych.

Etap pierwszy projektowania procesu technologicznego po otrzymaniu danych konstrukcyjnych dotyczy opracowania ramowego procesu technologicznego. Następnie ramowy proces technologiczny jest uzupełniany o dodatkowe dane dotyczące operacji technologicznych.

Etapy kolejne projektowania procesu technologicznego dotyczą doboru obrabiarki, narzędzi i parametrów obróbki dla określonych operacji technologicznych.

Etapem ostatnim jest generowanie procesu technologicznego w postaci karty

technologicznej i instrukcji na stanowiska obróbcze.

Wiedza technologiczna pochodzi z wielu źródeł. Dane z katalogów i baz danych można uzyskać w prosty sposób. Jednakże, jeśli chcemy pozyskać wiedzę, preferencje i doświadczenie technologa, należy wykorzystać bardziej zaawansowane narzędzia informatyczne, do których należy zaliczyć metody maszynowego uczenia, w tym metodę indukcji drzew decyzyjnych.

### 3. Metoda indukcji drzew decyzyjnych

Jedną z metod maszynowego uczenia jest metoda indukcji drzew decyzyjnych. Pozwala ona na przybliżenie funkcji klasyfikacyjnych o dyskretnych wartościach wejściowych odnoszących się do pewnych pojęć, klas decyzyjnych. Dzisiaj drzewa decyzyjne stanowią podstawową metodę indukcyjnego uczenia się maszyn [7]. Spowodowane jest to dużą efektywnością, możliwością prostej programowej implementacji, jak i intuicyjną czytelnością dla człowieka. Ta metoda pozyskiwania wiedzy opiera się na analizie przykładów, przy czym każdy przykład musi być opisany przez zestaw atrybutów, gdzie każdy atrybut może przyjmować różne wartości.

Drzewo decyzyjne jest to skierowany graf acykliczny, przy czym krawędzie takiego grafu nazywane są gałęziami, wierzchołki, z których wychodzi, co najmniej jedna krawędź, nazywane są węzłami, a pozostałe wierzchołki - liśćmi. Ponadto przyjmuje się, że w takim grafie istnieje tylko jedna ścieżka między różnymi wierzchołkami. Konstrukcję drzewa na podstawie zbioru przykładów, najprościej przedstawić w postaci algorytmu rekurencyjnego uruchamianego dla każdego węzła w drzewie. Pierwszym krokiem algorytmu jest podjęcie decyzji, czy rozpatrywany węzeł powinien stać się liściem końcowym drzewa wg kryterium stopu albo węzłem-rozgałęzieniem. Zaklasyfikowanie węzła jako liść końcowy drzewa spowoduje zakończenie algorytmu. Podjęcie decyzji, że dany węzeł staje się rozgałęzieniem spowoduje wybór atrybutu (z dostępnej puli atrybutów) wg kryterium wyboru atrybutu. Następnie na podstawie wartości, jakie przyjmuje wybrany argument, z zestawu przykładów tworzone są kolejne węzły drzewa. Algorytm uruchamiany jest rekurencyjnie. Odpowiednia technika wyboru argumentu ma kluczowy wpływ na wygląd drzewa decyzyjnego, gdyż właśnie od kolejności wyboru atrybutów zależy w głównej mierze głębokość i stopień rozbudowy drzewa. Wywołania rekurencyjne algorytmu tworzącego drzewo należy kiedyś zakończyć, odpowiada za to właśnie kryterium stopu. Określa ono, czy dany węzeł drzewa powinien być traktowany jako końcowy liść drzewa zawierający w swoim opisie etykietę klasy-decyzji. Natomiast kryterium wyboru atrybutu jest to w zasadzie najważniejsza część algorytmu, to od niej zależy kolejność wyboru atrybutów do zbudowania testu, na którego podstawie nastąpi w węźle podział zbioru przykładów. W znaczącym stopniu wpływa to na późniejszy wygląd drzewa. Można stwierdzić, że dobrym testem jest ten, którego użycie w węźle spowoduje skrócenie ścieżki prowadzącej przez ten węzeł do liści wskazujących klasę decyzyjną. Wybór odpowiedniego atrybutu ze zbioru atrybutów jest dokonywany dzięki wprowadzeniu systemu ocen. System ocen atrybutów opiera się na założeniu, iż najbardziej bezużytecznym atrybutem jest taki, w którym rozkład częstości występowania kolejnych klas-wyboru jest taki sam przed i po podziale zbioru danych przykładów według ocenianego atrybutu. Do tworzenia drzewa zostało wykorzystane kryterium wyboru na podstawie miary przyrostu informacji, które jest powiązane z pojęciem miary entropii wywodzącej się z teorii informacji [8,9,10].

W pracy zostały wykorzystane wybrane algorytmy do budowy drzew klasyfikacyjnych. Do budowy modeli doboru obrabiarki, narzędzi i parametrów obróbki zastosowano: ogólne

modele drzew klasyfikacyjnych (C&RT), ogólne modele CHAID, wzmacniane drzewa klasyfikacyjne oraz las losowy.

Klasyczny algorytm *C&RT* rozpropagowany został przez Breimana i innych [11,12, 13]. Tworzy on zawsze drzewa binarne, które czasami nie dadzą się w przejrzysty sposób przedstawić w postaci podsumowania.

Akronim *CHAID* oznacza Chi-squared Automatic Interaction Detector. Jest to jedna z najstarszych metod drzew klasyfikacyjnych, zaproponowana przez Kassa [14]. W odróżnieniu od algorytmu C&RT, CHAID nie buduje drzew binarnych (tzn. buduje drzewa, w których z węzłów mogą wychodzić więcej niż dwie gałęzie) i wykonuje to za pomocą stosunkowo prostego algorytmu, który nadaje się zwłaszcza do analizy dużych zbiorów danych. CHAID tworzy drzewa nie-binarne, które są zwykle "szersze". CHAID często daje w wyniku wiele węzłów wychodzących z jednego, co łatwo przedstawić w prostej tabeli wielozdzielczej o wielu kategoriach dla każdej zmiennej lub wymiaru tabeli [13].

*Wzmacniane drzewa* są implementacją metody stochastycznego gradientowego wzmacniania drzew [13, 15]. Na przestrzeni ostatnich kilku lat technika ta rozwinęła się jako jedna z najpotężniejszych metod predykcyjnego data mining. Główną ideą jest tworzenie ciągu (bardzo) prostych drzew, z których każde kolejne jest zbudowane do predykcji reszt generowanych przez poprzednie. Metoda buduje drzewa binarne, tzn. dzieli dane na dwie próby w każdym węźle podziału. Przypuśćmy, że mamy ograniczyć złożoność drzew do 3 węzłów: tzn. drzewo składa się z korzenia i dwóch potomków, czyli jest tylko jeden podział. W kolejnych krokach wzmacniania (algorytmu wzmacniania drzew) określany jest pojedynczy (najlepszy) podział danych i obliczane są odchyłki wartości obserwowanych od średnich (reszty w każdym podziale). Kolejne drzewo o trzech węzłach jest dopasowywane do tych reszt i wyznacza kolejny podział, który jest dopasowywany do tych reszt i wyznacza kolejny podział, przy którym wariancja reszt (czyli błąd) jest jeszcze mniejsza (dla danego wcześniej ciągu drzew). Taka procedura addytywnego rozwinięcia ważonego drzew daje w efekcie doskonale dopasowanie wartości przewidywanych do wartości obserwowanych, nawet jeśli sama natura relacji pomiędzy predyktorami a zmienną zależną jest bardzo złożona (np. nieliniowa). Tak więc metoda wzmacniania gradientowego - dopasowania ważonego rozwinięcia addytywnego prostych drzew - jest bardzo ogólnym i mocnym algorytmem uczenia maszyn.

*Las losowy* jest w pełni funkcjonalną realizacją algorytmu Losowy las opracowanego przez Breimana [13, 16]. Losowy las to zespół prostych drzew, przewidujących wartość zmiennej zależnej na podstawie zbioru zmiennych niezależnych (predyktorów). W przypadku zadań klasyfikacyjnych wynikiem jest przewidywana przynależność do klasy (o której informuje nas wartość jakościowej zmiennej zależnej). Losowy las składa się z pewnej liczby prostych drzew. Przewidywanie zespołu uzyskiwane jest przez głosowanie dla zadań klasyfikacyjnych. Zespół prostych drzew zazwyczaj daje zdecydowanie trafniejsze przewidywania niż pojedyncze, nawet bardzo złożone drzewo.

Pod względem trafności predykcji trudno polecać w sposób szczególny określoną metodę. Jest to nadal przedmiot badań. Z praktycznego punktu widzenia, najlepiej zastosować różne algorytmy, po czym porównać je i następnie wybrać model najlepszy pod względem błędu predykcji [13].

#### 4. CASE Study

Dane rzeczywiste pochodzą z firmy oferującej szeroki asortyment produktów. Główne obszary działalności firmy to:

- produkcja okrętowa (okna, iluminatory, wycieraczki),
- produkcja armatury przemysłowej,
- produkcja szkła,
- obróbka metali,
- wyroby dla górnictwa.

Na potrzeby pracy została przeanalizowana obróbka metali. Niektóre części wykonywane są jako elementy jednostkowe, inne natomiast można zaliczyć do produkcji mało lub średnioseryjnej. Podczas produkcji wyrobów wykonywane są podstawowe operacje technologiczne, takie jak: cięcie, cięcie laserem, frezowanie, toczenie, szlifowanie, gwintowanie i inne. W pracy przedstawiono modele dla procesu frezowania.

W celu opracowania modeli doboru wspomagających projektowanie procesu technologicznego ustalono następujące kryteria:

- kryteria doboru narzędzi obróbkowych do operacji technologicznych zawierają następujące czynniki: sposób obróbki, kształt obrabianych powierzchni, rodzaj i dokładność obróbki, wielkość produkcji, materiał obrabianego przedmiotu oraz typ obrabiarki,
- kryteria doboru obrabiarek do operacji technologicznych uwzględniają następujące czynniki: rodzaj i zakres obróbki, wymiary przedmiotu obróbki, dokładność obróbki, wydajność produkcyjną, przewidywane obciążenie oraz godzinowy koszt pracy obrabiarki,
- kryteria doboru parametrów obróbki, na które mają istotny wpływ następujące czynniki: materiał przedmiotu obrabianego, materiał ostrza narzędzi, obrabiarka, rodzaj obróbki oraz wymagania dotyczące jakości powierzchni.

#### 4.1. Dobór narzędzi

##### Przygotowanie danych

W celu przygotowania danych uczących do budowy drzew klasyfikacyjnych została wykonana analiza narzędzi firmy produkcyjnej, które podzielono na: wiertła, frezy, noże tokarskie oraz ściernice. Dla każdej operacji technologicznej dobór narzędzi został wykonany osobno.

Na podstawie danych o narzędziach oraz kryteriów doboru został przygotowany plik uczący. Przypadki klasyfikacyjne przygotowano w następujący sposób. Wyróżniono predyktory ilościowe i jakościowe, takie jak: rodzaj operacji (np. zgrubna, wykańczająca), rodzaj powierzchni obrabianej (rowek, obrys, płaszczyzna), rodzaj materiału obrabianego (np. 316L), chropowatość (np. 20), rodzaj (np. łączone), rodzaj mocowania freza (np. nasadowe), średnica freza [mm], kształt freza (walcowy), liczba ostrzy (np. 10), długość całkowita freza [mm], prędkość skrawania  $v_c$  [m/min], głębokość skrawania  $a_p$  [mm], posuw  $f$  [mm/min], koszt eksploatacji, szerokość frezowania  $a_e$  [mm]. Zmienną zależną jest symbol freza. Plik uczący zawiera 553 rekordy.

##### Modele doboru narzędzi

Eksperymenty zostały przeprowadzone dla wybranych operacji technologicznych. W pracy zostały szczegółowo przedstawione modele dla operacji frezowania. Wykonano modele w postaci drzew klasyfikacyjnych: C&RT, CHAID, wzmocnionych drzew oraz losowego lasu. Dla każdego typu drzew zmieniano parametry modeli.

Dla modelu C&RT ustawiano parametry klasyfikacji: koszty błędnych klasyfikacji, dobroć dopasowania oraz prawdopodobieństwo a priori. Kryterium stopu zawierało regułę stopu: przytnij wg wariancji oraz parametr minimalnej liczności przykładów w węźle.

Dla modelu CHAID ustawiano koszty błędnych klasyfikacji. Kryterium stopu dotyczyło minimalnej liczności przykładów w węźle.

Dla modelu wzmacnianych drzew klasyfikacyjnych ustawiano parametry klasyfikacji: koszty błędnych klasyfikacji oraz prawdopodobieństwo a priori. Kryterium stopu zawierało parametr minimalnej liczności przykładów w węźle.

Dla modelu losowego lasu ustawiano parametry klasyfikacji: koszty błędnych klasyfikacji oraz prawdopodobieństwo a priori. Kryterium stopu zawierało parametr minimalnej liczności przykładów w węźle.

Koszty błędnych klasyfikacji dotyczą rozkładu przykładów pomiędzy klasami. Minimalizacja kosztów odpowiada minimalizacji proporcji przypadków zaklasyfikowanych błędnie, gdy bierze się prawdopodobieństwa a priori proporcjonalne do wielkości klas, a koszty błędnej klasyfikacji równe dla każdej klasy [13].

Dobroć dopasowania dotyczy znalezienia takiego podziału dla każdej zmiennej predykcyjnej w celu znalezienia podziału, dla którego następuje największy wzrost dobroci dopasowania [13]. Jak określić poprawę dobroci dopasowania? Dostępne są trzy miary dobroci dopasowania. Miara Giniego zanieczyszczenia węzła osiąga wartość zero, gdy w danym węźle wystąpi tylko jedna klasa (przy prawdopodobieństwach a priori oszacowanych na podstawie wielkości klas oraz równych kosztach błędnej klasyfikacji miarę Giniego oblicza się jako sumę iloczynów wszystkich par proporcji klas w danym węźle; osiąga ona wartość maksymalną, gdy liczności klas w danym węźle są równe). Miara Giniego była preferowaną miarą dobroci dopasowania przez twórców programu C&RT [11]. Doskonałe dopasowanie oznacza doskonałą klasyfikację.

Prawdopodobieństwa a priori określają, na ile jest prawdopodobne, bez żadnej wstępnej wiedzy na temat wartości zmiennych predykcyjnych w modelu, że dany przypadek lub obiekt należy do danej klasy. Stosowane są w minimalizacji kosztów i mogą wpływać na klasyfikację przypadków lub obiektów. Minimalizacja kosztów odpowiada minimalizacji ogólnej proporcji przypadków zaklasyfikowanych błędnie, wtedy gdy prawdopodobieństwa a priori będą proporcjonalne do wielkości klas (a koszty błędnych klasyfikacji równe dla każdej klasy), ponieważ przewidywanie powinno być lepsze w przypadku większych klas, aby dać niższą ogólną stopę błędnych klasyfikacji [13].

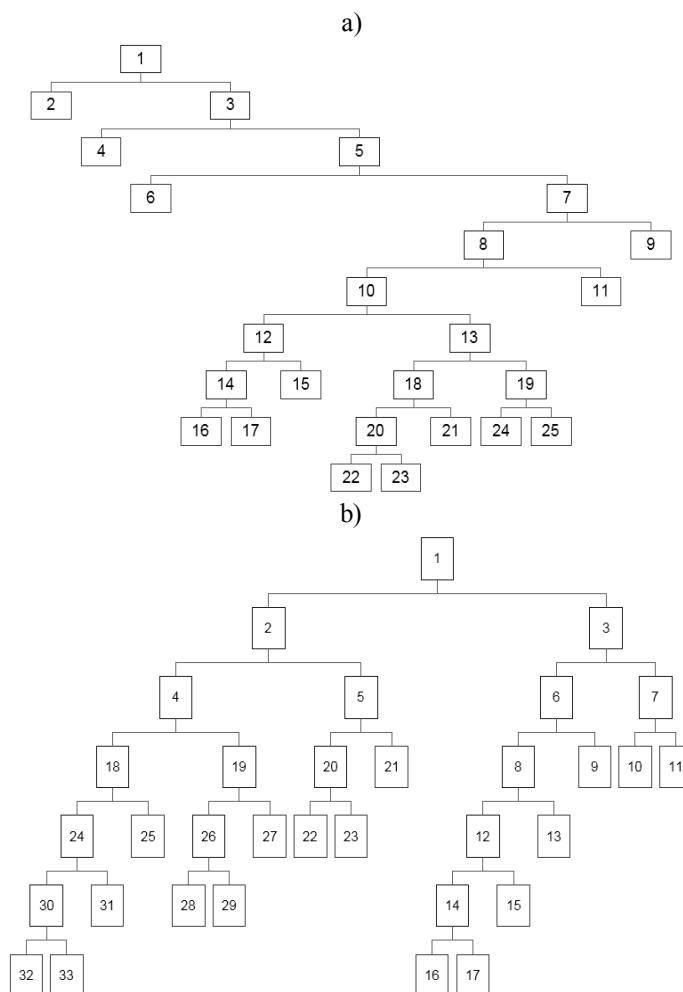
Na rysunku 1 pokazano przykładowe drzewa typu C&RT i CHAID. Drzewo klasyfikacyjne CHAID jest bardziej złożone. Analiza drzew wykazała inny dobór ważności predyktorów przez drzewa.

Macierze klasyfikacji dla wybranych drzew klasyfikacyjnych zostały pokazane na rysunku 2. W przypadku przewidywania klasy w stosunku do klasy obserwowanej bardzo dobrze zachowują się wzmacniane drzewa oraz las losowy. Trochę gorzej drzewo C&RT i CHAID.

Do oceny modeli zastosowano sprawdzian krzyżowy na podstawie próby testowej oraz 10-krotny sprawdzian krzyżowy. W tabeli 1 porównane zostały koszt i ocena ryzyka oraz błąd standardowy wygenerowanych drzew klasyfikacyjnych.

Tab. 1. Porównanie klasyfikatorów dla doboru freza

Typ klasyfikatora	Ocena klasyfikatora	
	Koszt/ocena ryzyka	Standardowy błąd
C&RT	0,065217	0,010509
CHAID	0,201087	0,017060
Wzmacniane drzewa	0,000000	0,000000
Losowy las	0,031746	0,012753



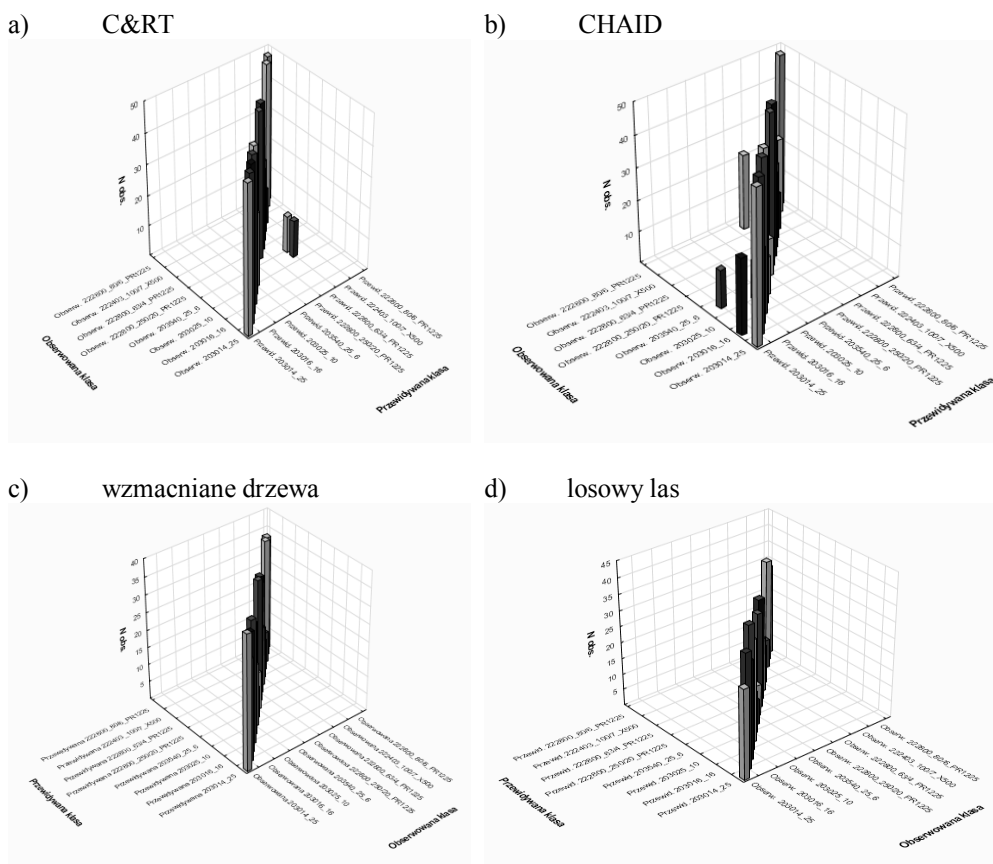
Rys. 1. Układ drzewa dla symbolu freza: a) C&RT, liczba węzłów dzielonych: 12, liczba węzłów końcowych: 13, b) CHAID, liczba węzłów dzielonych: 16, liczba węzłów końcowych: 17

Najlepszymi modelami okazały się wzmocnione drzewa oraz losowy las. Odpowiednio wartości poprawnej klasyfikacji wyniosły 100% i 97,64%.

## 4.2. Dobór obrabiarki

### Przygotowanie danych

W celu przygotowania danych uczących do budowy drzew klasyfikacyjnych została wykonana analiza parku maszynowego firmy produkcyjnej, w którym wyróżniono obrabiarki CNC: frezarki, wiertarko-frezarki, szlifierki oraz tokarki. Dla każdej operacji technologicznej dobór obrabiarki został wykonany osobno.



Rys. 2. Macierz klasyfikacji dla wybranych drzew klasyfikacyjnych

Na podstawie danych o obrabiarkach oraz kryteriów doboru został przygotowany plik uczący. Przypadki klasyfikacyjne przygotowano w następujący sposób. Wyróżniono predyktory ilościowe i jakościowe, takie jak: rodzaj operacji (np. zgrubna, wykańczająca), X - długość wyrobu [mm], Y - szerokość wyrobu/średnica [mm], Z - wysokość wyrobu/średnica [mm], X - wielkość przestrzeni roboczej [mm], Y - wielkość przestrzeni roboczej [mm], Z - wielkość przestrzeni roboczej [mm], max. średnica narzędzia [mm], długość narzędzia [mm], koszt eksploatacji obrabiarki [PLN/h], min. prędkość obrotowa [obr/min], max. prędkość obrotowa [obr/min], max. posuw roboczy  $f$  [mm/min] oraz moc obrabiarki (moc na silniku wrzeciona) [KW]. Zmienną zależną jest symbol obrabiarki. Plik uczący zawiera 521 rekordów.

#### Modele doboru obrabiarki

Modele doboru obrabiarki budowane były w podobny sposób jak modele doboru narzędzi. Wykonano modele w postaci drzew klasyfikacyjnych: C&RT, CHAID, wzmocnianych drzew oraz losowego lasu. W przypadku doboru obrabiarki również najlepsze okazały się wzmocnione drzewa oraz losowy las.



### 4.3. Dobór parametrów obróbki

#### Przygotowanie danych

W celu przygotowania danych uczących do budowy drzew klasyfikacyjnych została wykonana analiza procesów technologicznych pod kątem doboru parametrów obróbki dla określonej obrabiarki i narzędzi w firmie produkcyjnej. Dla każdej operacji technologicznej dobór parametrów obróbki został wykonany osobno.

Na podstawie danych o parametrach obróbki oraz kryteriów doboru został przygotowany plik uczący. Przypadki klasyfikacyjne przygotowano w następujący sposób. Wyróżniono predyktory ilościowe i jakościowe, takie jak: rodzaj operacji (np. zgrubna, wykańczająca), rodzaj materiału obrabianego (np. 316L), symbol wybranego narzędzia, chropowatość (np. 20), głębokość skrawania  $a_p$  [mm], szerokość frezowania  $a_e$  [mm], docelowa głębokość [mm], symbol obrabiarki. Zmienną zalezną jest zestaw parametrów obróbki, które należy ustawić na obrabiarce: posuw  $f$  [mm/min], prędkość skrawania [m/min], czas trwania [min] oraz dodatkowo trwałość ostrza [min]. Plik uczący zawiera 617 rekordów.

#### Modele doboru parametrów obróbki

Modele doboru parametrów obróbki budowane były w podobny sposób jak modele doboru narzędzi. Wykonano modele w postaci drzew klasyfikacyjnych: C&RT, CHAID, wzmacnianych drzew oraz losowego lasu. W przypadku doboru parametrów obróbki również najlepsze okazały się wzmacniane drzewa oraz losowy las.

### 5. Wnioski

W przypadku wspomaganie projektowania procesu technologicznego ze względu na dużą liczbę danych wejściowych w postaci symbolicznej właściwe jest zastosowanie metody indukcji drzew decyzyjnych, jako metody klasyfikacji. Drzewa decyzyjne wykazują bardzo dobre właściwości klasyfikacji. Generowanie reguł na podstawie drzew decyzyjnych umożliwia ich zwarty zapis i znacznie skraca czas potrzebny do wnioskowania.

Przeprowadzone badania wykazały użyteczność drzew klasyfikacyjnych oraz ich dużą skuteczność do wspomaganie projektowania procesu technologicznego. Drzewa klasyfikacyjne jako bardzo dobre algorytmy eksploracji danych dały duże możliwości korzystania z danych zawartych w technologicznych bazach danych.

Porównanie ogólnych modeli drzew klasyfikacyjnych (C&RT), ogólnych modeli CHAID, wzmacnianych drzew klasyfikacyjnych oraz losowego lasu dało również ciekawe wnioski badawcze. Najlepszymi modelami okazały się wzmacniane drzewa oraz losowy las. Wymagają one co prawda trochę więcej czasu na zbudowanie drzew, ale nie są to duże różnice w stosunku do drzew typu C&RT i CHAID.

Zastosowanie metody maszynowego uczenia okazało się bardzo cennym narzędziem do pozyskania wiedzy technologicznej. Wynika to z faktu, że często technolog nie potrafi wyjaśnić swojej decyzji, z której można utworzyć klasyczne reguły występujące w systemach ekspertowych. Dlatego automatyczne tworzenie reguł z przykładów już zaprojektowanych procesów technologicznych, zweryfikowanych podczas produkcji wyrobów jest bardzo dobrym rozwiązaniem odkrywania wiedzy wynikającej z doświadczenia technologów.

System pozyskiwania wiedzy technologicznej, w którym zawarte są modele klasyfikacji

może wspomagać mniej doświadczonych technologów w trakcie projektowania procesów technologicznych.

Jest to szczególnie istotne w sytuacjach, gdy wiedza taka nie jest dostępna, bądź jest trudna do sformalizowania, niepełna i niepewna.

*Praca powstała w ramach badań statutowych pt. „Techniki obliczeniowe w optymalizacji i modelowaniu problemów technicznych i środowiskowych” realizowanych w Instytucie Mechaniki i Informatyki Stosowanej na Uniwersytecie Kazimierza Wielkiego w Bydgoszczy.*

## Literatura

1. Feld M.: Podstawy projektowania procesów technologicznych typowych części maszyn, WNT, Warszawa, 2003.
2. Szatkowski K.: Przygotowanie produkcji. PWN, Warszawa, 2008.
3. Chlebus E.: Technika komputerowa CAx w inżynierii produkcji. WNT, 2000.
4. Weiss Z.: Techniki komputerowe w przedsiębiorstwie. Wydawnictwo Politechniki Poznańskiej, Poznań, 2002.
5. Knosala R. i in.: Zastosowania metod sztucznej inteligencji w inżynierii produkcji, WNT, Warszawa, 2002.
6. Rojek I.: Classifier Models in Intelligent CAPP Systems. Man-Machine Interactions, seria Advances in Intelligent and Soft Computing, Springer-Verlag, 2009, 311-319.
7. Krawiec K. Stefanowski J.: Uczenie maszynowe i sieci neuronowe. Wydawnictwo Politechniki Poznańskiej, Poznań, 2004.
8. Cichosz P.: Systemy uczące się. WNT, Warszawa, 2000.
9. Larose D. T.: Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych. PWN, Warszawa, 2006.
10. Russell S. J., Norvig P.: Artificial intelligence: a modern approach. Prentice Hall, New Jersey, 2009.
11. Breiman L., Friedman J. H., Stone C. J., Olshen R. A.: Classification and regression trees. Wadsworth, Belmont (CA), 1984.
12. Ripley B. D.: Pattern recognition and neural networks. Cambridge University Press, 1996.
13. StatSoft Statistica Internetowy podręcznik statystyki, [http://www.statsoft.pl/textbook/stathome\\_stat.html?http%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstclatre.html](http://www.statsoft.pl/textbook/stathome_stat.html?http%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstclatre.html), 10.01.2015.
14. Kass G. V.: An exploratory technique for investigating large quantities of categorical data. Applied Statistics, 29, 1980, 119-127.
15. Hastie T., Tibshirani R., Friedman J. H.: The elements of statistical learning: data mining, inference, and prediction. Springer series in statistics. World Publishing Corporation Beijing, 2001.
16. Breiman L.: Random forests. Machine Learning, 45, Springer-Verlag, 2001, 5-32.

Dr inż. Izabela ROJEK  
Instytut Mechaniki i Informatyki Stosowanej  
Uniwersytet Kazimierza Wielkiego  
85-064 Bydgoszcz, ul. Chodkiewicza 30  
tel./fax: (0-52) 325 76 30/ (0-52) 325 76 44  
e-mail: izarojek@ukw.edu.pl