IBM SPSS Decision Trees 29

Uwaga

Przed użyciem tych informacji i produktu, którego one dotyczą, przeczytaj informacje znajdujące się w sekcji “Uwagi” na stronie 21.

Informacje o produkcie

Niniejsze wydanie dotyczy wersji 29, wydania 0, modyfikacji 1 produktu IBM® SPSS Statistics oraz wszystkich kolejnych wydań i modyfikacji, dopóki nie zostanie to określone inaczej w nowych wydaniach.

Spis treści

Rozdział 1. Drzewa decyzyjne

Następujące funkcje dotyczące drzew decyzyjnych są dostępne w module SPSS Statistics Professional Edition oraz w module Drzewa decyzyjne.

Tworzenie drzew decyzyjnych

Procedura drzew decyzyjnych tworzy model klasyfikacji oparty na drzewie. Klasyfikuje obserwacje w grupy lub przewiduje wartości zależnej (przewidywanej) zmiennej w oparciu o wartości niezależnych zmiennych (predyktorów). Ta procedura udostępnia narzędzia walidacyjne przeznaczone do analizy eksploracyjnej lub potwierdzającej.

Ta procedura może być używana na potrzeby:

Segmentacji. Identyfikacja osób, które mogą należeć do konkretnej grupy.

Warstwowania. Przypisywanie obserwacji do jednej z kilku kategorii, takich jak grupy wysokiego, średniego i niskiego ryzyka.

Predykcji. Tworzenie reguł i używanie ich w celu przewidywania zdarzeń w przyszłości, takich jak prawdopodobieństwo tego, że ktoś nie będzie spłacał pożyczki, albo potencjalna wartość samochodu lub domu przy odsprzedaży.

Redukcja danych i filtrowanie zmiennych. Wybór użytecznego podzbioru predyktorów z dużego zestawu zmiennych w celu opracowania formalnego modelu parametrycznego.

Identyfikacja interakcji. Identyfikacja zależności, które dotyczą tylko konkretnych grup, a następnie określenie ich w formalnym modelu parametrycznym.

Scalanie kategorii i dyskretyzacja zmiennych ciągłych. Ponowne kodowanie kategorii predyktorów grup i zmiennych ciągłych z minimalnymi stratami informacji.

Przykład. Bank chce skategoryzować osoby składające wnioski o kredyty na podstawie tego, czy stanowią istotne ryzyko kredytowe, czy nie stanowią. Na podstawie różnych czynników, w tym na podstawie znanych rang kredytów dawnych klientów można opracować model, aby przewidzieć prawdopodobieństwo tego, że przyszli klienci nie będą spłacać pożyczek.

Analiza oparta o drzewo zapewnia kilka atrakcyjnych funkcji:

- Umożliwia identyfikację jednorodnych grup z wysokim lub niskim ryzykiem.
- Ułatwia tworzenie reguł przeznaczonych do wykonywania predykcji na temat poszczególnych obserwacji.

Zagadnienia dotyczące danych

Dane. Zmienne zależne i niezależne mogą być następujące:

- *Nominalny*. Zmienna może być traktowana jako nominalna, gdy jej wartości reprezentują kategorie bez wewnętrznego rankingu (na przykład dział przedsiębiorstwa, w którym pracuje pracownik). Przykładami zmiennych nominalnych są: region, kod pocztowy lub wyznanie.
- *Porządkowy*. Zmienna może być traktowana jako porządkowa, gdy jej wartości reprezentują kategorie z jakimś nieodłącznym rangą (na przykład poziomy zadowolenia z usługi z bardzo niezadowolonego do bardzo zadowolonego). Przykładami zmiennych porządkowych mogą być oceny opinii reprezentujące stopień satysfakcji lub przekonania oraz oceny preferencji.
- *scale*. Zmienna może być traktowana jako skala (ilościowa), gdy jej wartości reprezentują uporządkowane kategorie z miarodajnym pomiarem, tak aby porównania odległości między wartościami były odpowiednie. Przykładami zmiennych ilościowych mogą być: wiek w latach lub przychód w tysiącach złotych.

Wagi liczebności Jeżeli obowiązuje ważenie, wagi ułamkowe są zaokrąglane do najbliższej wartości całkowitej; dlatego do obserwacji z wartością wagi niższą niż 0,5 przypisywane są wagi 0 i dlatego są wykluczane z analizy.

Założenia. Ta procedura zakłada, że odpowiedni poziom pomiaru został przypisany do wszystkich zmiennych analizy, a niektóre funkcje zakładają, że wszystkie wartości zmiennej zależnej uwzględnionej w analizie mają zdefiniowane etykiety wartości.

- **Poziom pomiaru.** Poziom pomiaru wpływa na obliczenia w drzewie; dlatego do wszystkich zmiennych powinien być przypisany odpowiedni poziom pomiaru. Domyślnie obowiązuje założenie, że zmienne numeryczne są ilościowe, a zmienne łańcuchowe są nominalne, co niekoniecznie dokładnie odzwierciedla rzeczywisty poziom pomiaru. Ikona obok każdej zmiennej na liście zmiennych określa jej rodzaj.

Tabela 1. Ikony poziomu pomiaru

Ikona	poziom pomiaru
	Skala
	Nominalny
	Porządkowy

Można tymczasowo zmienić poziom pomiaru dla zmiennej, klikając prawym przyciskiem myszy zmienną na liście zmiennych źródłowych i wybierając poziom pomiaru z menu kontekstowego.

- **Etykiety wartości.** W interfejsie okna dialogowego dla tej procedury obowiązuje założenie, że etykiety wartości są zdefiniowane dla wszystkich wartości (bez braków danych) zależnej zmiennej jakościowej (nominalnej, porządkowej) albo nie są zdefiniowane dla żadnych z tych wartości. Niektóre funkcje są niedostępne, chyba że etykiety wartości istnieją dla co najmniej dwóch wartości bez braków danych zależnej zmiennej jakościowej. Jeśli co najmniej dwie wartości bez braków danych mają zdefiniowane etykiety wartości, wszelkie obserwacje z innymi wartościami, które nie mają etykiet, są wykluczane z analizy.

Uzyskiwanie drzew decyzyjnych

1. Wybierz z menu następujące opcje:

   Analiza > **Klasyfikacja** > **Drzewo...**
2. Wybierz zmienną zależną.
3. Wybierz co najmniej jedną zmienną niezależną.
4. Wybierz metodę budowy.

Opcjonalnie możesz wykonać następujące czynności:

- Zmień poziom pomiaru dla dowolnej zmiennej z listy źródeł.
- Wymuś wprowadzenie do modelu pierwszej zmiennej z listy niezależnych zmiennych jako pierwszej zmiennej podzielonej.
- Wybierz zmienną wpływu, która definiuje stopień wpływu obserwacji na proces wzrostu drzewa. Obserwacje z mniejszą wartością wpływu mają mniejszy wpływ, obserwacje z większą wartością wpływu mają większy wpływ. Zmienna wpływu musi być dodatnia.
- Przeprowadź walidację drzewa.
- Dostosuj kryterium wzrostu drzewa.
- Zapisz numery węzłów końcowych, wartości przewidywane oraz przewidywane prawdopodobieństwa jako zmienne.

- Zapisz model w formacie XML (PMML).

Zmienne z nieznanym poziomem pomiaru

Alert poziomu pomiaru wyświetla się, gdy poziom pomiaru dla jednej lub większej ilości zmiennych w zbiorze danych jest nieznany. Ponieważ poziom pomiaru wpływa na wyliczenie wyników dla tej procedury, wszystkie zmienne muszą mieć zdefiniowany poziom pomiaru.

Skanowanie danych. Odczytuje dane w aktywnym zbiorze danych i przypisuje domyślny poziom pomiaru do wszystkich zmiennych, które mają aktualnie nieznany poziom pomiaru. Jeśli zbiór danych jest duży, może to zająć trochę czasu.

Przypisz ręcznie. Otwiera okno dialogowe, które zestawia wszystkie zmienne z nieznanym poziomem pomiaru. Można użyć tego okna dialogowego do przypisania poziomu pomiaru do tych zmiennych. Można również przypisać poziom pomiaru w Widoku zmiennych Edytora danych.

Ponieważ poziom pomiaru jest ważny dla tej procedury, nie można wejść do tego okna dialogowego w celu uruchomienia tej procedury, dopóki wszystkie zmienne nie będą miały zdefiniowanego poziomu pomiaru.

Zmienianie poziomu pomiaru

1. Kliknij prawym przyciskiem myszy zmienną na liście źródłowej.
2. Z menu kontekstowego wybierz poziom pomiaru.

Spowoduje to tymczasową zmianę poziomu pomiaru i umożliwi użycie go w procedurze drzewa decyzyjnego.

Metody budowy

Dostępne metody budowy są następujące:

CHAID. Chi-kwadrat-automatyczne wykrywanie interakcji Chi-kwadrat. Na każdym etapie funkcja CHAID wybiera niezależną zmienną (predyktor), która ma najsilniejszą interakcję z niezależną zmienną. Jeśli w odniesieniu do zmiennej niezależnej kategorie nie różnią się znacznie od siebie, to Kategorie wszystkich predyktorów są połączone.

Wyczerpujący CHAID. Zmodyfikowany CHAID badający wszystkie możliwe podziały wszystkich predyktorów.

CRT. drzewa klasyfikacji i regresji. Ze względu na zmienną zależną CRT dzieli dane na jak najbardziej jednorodne segmenty. Węzeł końcowy, w którym wszystkie obserwacje dla zmiennej zależnej mają identyczne wartości jest jednorodnym „czystym" węzłem.

QUEST. Szybkie, Nietendencyjnie, Efektywne Drzewo Statystyczne. Metoda ta jest szybka i jednocześnie zapobiega odchyleniom innych metod na rzecz predyktorów z wieloma kategoriami. Metody SNWDS można używać, tylko gdy zależna zmienna jest normalna.

W przypadku każdej metody istnieją zalety i ograniczenia, a wśród nich następujące:

Tabela 2. Cechy metody budowy

Opcja	**CHAID***	**CRT**	**QUEST**
Oparta o Chi-kwadrat**	X		
Zastępowanie zmiennych niezależnych (predyktorów)		X	X
Obcinanie drzewa		X	X
Wielokrotny podział węzłów	X		
Binarny podział węzłów		X	X
Zmienne wpływu	X	X	
Prawdopodobieństwa a priori		X	X

Tabela 2. Cechy metody budowy (kontynuacja)			
Opcja	**CHAID***	**CRT**	**QUEST**
Koszty błędnej klasyfikacji	X	X	X
Szybkie obliczenie	X		X

*Obejmuje wyczerpujący CHAID

**QUEST używa także miary chi-kwadrat względem nominalnych zmiennych niezależnych.

Wybór kategorii

W przypadku jakościowych (nominalnych, porządkowych) zmiennych zależnych można:

- Kontrolować kategorie uwzględnione w analizie.
- Wskazywać badane kategorie zmiennych przewidywanych.

Uwzględnianie/wykluczanie kategorii

Analizę można ograniczyć do konkretnych kategorii zmiennej zależnej.

- Obserwacje z wartościami zmiennej zależnej, które znajdują się na liście Wyklucz brakujące wartości, nie są uwzględniane w analizie.
- W przypadku nominalnych zmiennych zależnych można w analizie uwzględnić także kategorie będące brakami danych użytkownika. (Domyślnie kategorie będące brakami danych użytkownika są wyświetlane na liście Wyklucz brakujące wartości).

Kategorie docelowe

Wybrane (zaznaczone) kategorie są traktowane jako kategorie o podstawowym znaczeniu w analizie. Jeśli na przykład interesujesz się przede wszystkim znalezieniem tych osób, które najprawdopodobniej nie będą spłacać pożyczki, możesz — jako kategorię docelową — wybrać kategorię oceny zdolności kredytowej „Zła”.

- Nie istnieje domyślna kategoria docelowa. Jeśli nie została wybrana żadna kategoria, niedostępne są niektóre opcje reguł klasyfikacji oraz opcje danych wyjściowych powiązane z korzyściami.
- W przypadku wyboru wielu kategorii można dla każdej kategorii docelowej wygenerować osobne tabele i wykresy korzyści.
- Wyznaczenie co najmniej jednej kategorii jako docelowej nie ma wpływu na model drzewa, ocenę ryzyka ani wyniki błędnej klasyfikacji.

Kategorie i etykiety wartości

To okno dialogowe wymaga zdefiniowanych etykiet wartości zmiennej zależnej. Aby było dostępne, co najmniej dwie wartości jakościowej zmiennej zależnej muszą mieć zdefiniowane etykiety wartości. .

Uwzględnianie/wykluczanie kategorii oraz wybór kategorii docelowych

1. W głównym oknie dialogowym Drzewo klasyfikacyjne wybierz jakościową zmienną zależną (nominalną, porządkową) z co najmniej dwiema zdefiniowanymi etykietami wartości.
2. Kliknij opcję **Kategorie**.

Sprawdzanie

Walidacja umożliwia ocenę tego, jak dobrze struktura drzewa uogólnia większą populację. Dostępne są dwie metody walidacji: walidacja krzyżowa oraz walidacja z podziałem próby.

Walidacja krzyżowa

Walidacja krzyżowa dzieli próbę na kilka podprób — **tzw. krotności**. Następnie generowane są modele drzewa, wyłączając kolejno dane z każdej podpróby. Pierwsze drzewo jest oparte na wszystkich obserwacjach z wyjątkiem tych w pierwszej krotności próby; drugie drzewo jest oparte na wszystkich

obserwacjach z wyjątkiem drugiej krotności próby itd. Dla każdego drzewa szacowane jest ryzyko błędnej klasyfikacji z zastosowaniem drzewa na podpróbie wyłączonej podczas generowania tego drzewa.

- Można określić maksymalnie 25 krotności próby. Im wyższa jest wartość, tym mniejsza liczba obserwacji wykluczanych z każdego drzewa modelu.
- Walidacja krzyżowa zwraca pojedynczy, ostateczny model drzewa. Ocena ryzyka z walidacji krzyżowej dla ostatecznego drzewa jest obliczana jako średnia ryzyk dla wszystkich drzew.

Walidacja z podziałem próby

W przypadku walidacji z podziałem próby model jest generowany z użyciem próby uczącej i testowany na próby wstrzymanej.

- Użytkownik może określić rozmiar próby uczącej wyrażony jako procent całkowitego rozmiaru próby albo może określić zmienną, która podzieli próbę na próby uczące i testujące.
- Jeśli w celu zdefiniowania prób uczących i testujących zostanie użyta zmienna, obserwacje z wartością 1 dla zmiennej zostaną przypisane do próby uczącej, a wszystkie pozostałe obserwacje zostaną przypisane do próby testującej. Zmienna nie może być zmienną zależną, zmienną ważącą, zmienną wpływu ani wymuszoną zmienną niezależną.
- Wyniki można wyświetlić dla prób uczących i testujących albo tylko dla próby testującej.
- Walidacja z podziałem próby powinna być stosowana ostrożnie w przypadku małych plików danych (pliki danych z małą liczbą obserwacji). Małe próby uczące mogą zwracać modele niskiej jakości, ponieważ niektóre kategorie mogą zawierać niewystarczającą ilość obserwacji do zapewnienia odpowiedniego wzrostu drzewa.

Aby zwalidować drzewo decyzyjne

1. W głównym oknie dialogowym Drzewa decyzyjne kliknij opcję **Walidacja**.
2. Wybierz opcję **Walidacja krzyżowa** lub **Walidacja z podziałem próby**.

Uwaga: W obu metodach walidacji następuje losowe przydzielanie obserwacji do grup prób. Jeśli wymagana jest możliwość reprodukcji dokładnie tych samych wyników w kolejnej analizie, należy ustawić wartość startową generatora liczb losowych (menu Przekształcenia, Generatory liczb losowych) przed uruchomieniem analizy po raz pierwszy, a następnie należy zresetować wartość startową do tej wartości przy następnej analizie. .

Kryterium wzrostu drzewa

To, jakie kryteria wzrostu są dostępne, może zależeć od metody wzrostu, poziomu pomiaru zmiennej zależnej lub kombinacji tych czynników.

Ograniczenia wzrostu

Karta Ograniczenia wzrostu umożliwia ograniczenie liczby poziomów w drzewie i kontrolowanie minimalnej liczby obserwacji dla węzłów nadrzędnych i podrzędnych.

Maksymalna głębokość drzewa. Kontroluje maksymalną liczbę poziomów wzrostu poniżej węzła głównego. Ustawienie **Automatyczne** ogranicza drzewo do trzech poziomów poniżej węzła głównego dla metod CHAID i Wyczerpujący CHAID oraz do pięciu poziomów dla metod CRT i QUEST.

Minimalna liczba obserwacji. Kontroluje minimalne liczby obserwacji dla węzłów. Węzły, które nie spełniają tych kryteriów, nie będą dzielone.

- W przypadku zwiększenia wartości minimum występuje tendencja do generowania drzew z mniejszą liczbą węzłów.
- W przypadku zmniejszenia wartości minimum powstają drzewa z większą liczbą węzłów.

W przypadku plików danych z małą liczbą obserwacji domyślna wartość 100 obserwacji dla węzłów nadrzędnych i 50 obserwacji dla węzłów podrzędnych może czasami powodować powstawanie drzew bez węzłów poniżej węzła głównego; w takim przypadku obniżenie wartości minimum może zapewnić bardziej użyteczne wyniki.

Określanie ograniczeń wzrostu

1. W głównym oknie dialogowym Drzewo decyzyjne kliknij opcję **Kryteria**.
2. Kliknij kartę **Ograniczenia wzrostu**.

Kryteria CHAID

W przypadku metod CHAID i Wyczerpujący CHAID można kontrolować:

Poziom istotności. Można kontrolować wartość istotności na potrzeby podziału węzłów i scalania kategorii. W przypadku obu tych kryteriów domyślnym poziomem istotności jest 0,05.

- W przypadku węzłów podziału wartość musi być większa od 0 i mniejsza niż 1. Niższe wartości wykazują tendencję do tworzenia drzew o mniejszej liczbie węzłów.
- W przypadku scalania kategorii wartość ta musi być większa od 0 i mniejsza lub równa 1. Aby zapobiec scalaniu kategorii, należy określić wartość 1. W przypadku ilościowej zmiennej niezależnej oznacza to, że liczba kategorii dla zmiennej w drzewie końcowym jest określoną liczbą przedziałów (wartość domyślna to 10). Więcej informacji zawiera temat “Przedziały zmiennych ilościowych dla analizy CHAID” na stronie 6.

Statystyki chi-kwadrat. W przypadku porządkowych zmiennych zależnych wartość chi-kwadrat do ustalenia podziału węzłów i scalania kategorii jest obliczana z użyciem metody ilorazu wiarygodności. W przypadku nominalnych zmiennych zależnych można wybrać metodę:

- **Pearsona.** Ta metoda skraca czas obliczeń, lecz należy zachować ostrożność w przypadku stosowania jej do niewielkich prób. Jest to metoda domyślna.
- **Iloraz wiarygodności.** Ta metoda jest bardziej odporna niż Pearsona, ale jej obliczenia trwają dłużej. Jest to metoda preferowana w przypadku niewielkich prób.

Estymacja modelu. W przypadku nominalnych i porządkowych zmiennych zależnych można określić:

- **Maksymalna liczba iteracji.** Wartość domyślna to: 100. Jeśli drzewo przestaje rosnąć z powodu osiągnięcia maksymalnej liczby iteracji, można powiększyć maksimum albo zmienić co najmniej jedno kryterium, które kontroluje wzrost drzewa.
- **Minimalna zmiana w oczekiwanych częstościach komórek.** Wartość musi być liczbą większą od 0 i mniejszą od 1. Domyślną wartością jest 0,05. Niższe wartości wykazują tendencję do tworzenia drzew o mniejszej liczbie węzłów.

Skoryguj wartości istotności metodą Bonferroniego. W przypadku porównań wielokrotnych wartości istotności dla kryteriów scalania i dzielenia są dostosowywane z użyciem metody Bonferroniego. Jest to wartość domyślna.

Zezwalaj na ponowny podział połączonych kategorii w węzłach. Jeśli scalanie kategorii nie zostanie jawnie zablokowane, ta procedura podejmie próbę scalenia kategorii zmiennych niezależnych (predyktorów) w celu uzyskania najprostszego modelu, który opisuje model. Ta opcja umożliwia procedurze ponowny podział scalonych kategorii, jeśli w ten sposób możliwe będzie uzyskanie lepszego rozwiązania.

Aby określić kryteria CHAID

1. W głównym oknie dialogowym Drzewo klasyfikacyjne wybierz **CHAID** albo **Wyczerpujący CHAID** jako metodę wzrostu.
2. Kliknij opcję **Kryteria**.
3. Kliknij kartę **CHAID**.

Przedziały zmiennych ilościowych dla analizy CHAID

W analizie CHAID ilościowe zmienne niezależne (predyktor) są zawsze przed analizą dzielone na grupy dyskretne (na przykład 0-10, 11-20, 21-30 itd.). Można określić początkową/maksymalną liczbą grup (jednak procedura może scalić przyległe grupy po początkowym podziale):

- **Ustalona liczba przedziałów.** Wszystkie ilościowe zmienne niezależne są początkowo dzielone na tę samą liczbę przedziałów. Wartością domyślną jest 10.
- **Użytkownika.** Każda ilościowa zmienna niezależna jest początkowo dzielona na określoną dla niej liczbę przedziałów.

Aby określić przedziały dla ilościowych zmiennych niezależnych

1. W głównym oknie dialogowym Drzewo klasyfikacyjne wybierz co najmniej jedną ilościową zmienną niezależną.
2. Jako metodę wzrostu wybierz **CHAID** lub **Wyczerpujący CHAID**.
3. Kliknij opcję **Kryteria**.
4. Kliknij kartę **Przedziały**.

W analizie CRT i QUEST wszystkie podziały są binarne, a niezależne zmienne ilościowe i porządkowe są traktowane tak samo. Nie można zatem określić liczby przedziałów dla ilościowych zmiennych niezależnych.

Kryteria CRT

Metoda wzrostu CRT ukierunkowana jest na uzyskanie maksymalnej jednorodności w obrębie każdego węzła. Stopień, w jakim węzeł nie odzwierciedla jednorodnego podzbioru obserwacji, jest miarą **zanieczyszczenia**. Na przykład węzeł końcowy, w którym wszystkie obserwacje mają tę samą wartość zmiennej zależnej, jest jednorodny i nie wymaga dalszego podziału, ponieważ jest „czysty”.

Można wybrać metodę pomiaru zanieczyszczenia i minimalny spadek zanieczyszczenia wymagany do podziału węzłów.

Miary zanieczyszczenia. W przypadku ilościowych zmiennych zależnych zanieczyszczenie wyrażone jest jako odchylenie liczone metodą najmniejszych kwadratów (LSD). Obliczane jest jako wariancje wewnątrz węzła i korygowane z uwzględnieniem wag częstości lub wartości wpływów.

W przypadku jakościowych (nominalnych, porządkowych) zmiennych zależnych można wybrać miarę zanieczyszczenia:

- **Gini.** Podziały są dokonywane w sposób, który maksymalizuje jednorodność węzłów podrzędnych ze względu na wartość zmiennej zależnej. Metoda Gini bazuje na kwadratach prawdopodobieństwa przynależności do każdej kategorii zmiennej zależnej. Miara osiąga minimum (zero), gdy wszystkie obserwacje w węźle należą do jednej kategorii. Jest to miara domyślna.
- **Twoing.** Kategorie zmiennej zależnej są grupowane w dwie podklasy. Podziały są dokonywane w oparciu o najlepszy separator przydzielający do dwóch grup.
- **Porządkowy Twoing.** Podobna do miary Twoing, ale tylko sąsiednie kategorie mogą być grupowane. Ta miara jest dostępna tylko dla porządkowych zmiennych zależnych.

Minimalna zmiana w ulepszeniu. Jest to minimalny spadek zanieczyszczenia wymagany, by węzeł został podzielony. Domyślną wartością jest 0,0001. Wyższe wartości wykazują tendencję do tworzenia drzew o mniejszej liczbie węzłów.

Aby określić kryteria CRT

1. Jako metodę wzrostu wybierz **CRT**.
2. Kliknij opcję **Kryteria**.
3. Kliknij kartę **CRT**.

Kryteria QUEST

W metodzie QUEST można określić poziom istotności decydujący o podziale węzłów. Zmienna niezależna nie może być wykorzystana do podziału węzła, jeśli jej wartość istotności jest mniejsza lub równa poziomowi zdefiniowanemu przez użytkownika. Wartość musi być liczbą większą od 0 i mniejszą od 1. Domyślną wartością jest 0,05. Przy mniejszych wartościach pojawi się tendencja do wykluczania większej liczby zmiennych niezależnych z ostatecznego modelu.

Aby określić kryteria QUEST

1. W głównym oknie dialogowym Drzewo klasyfikacyjne wybierz nominalną zmienną zależną.
2. Jako metodę wzrostu wybierz **QUEST**.
3. Kliknij opcję **Kryteria**.
4. Kliknij kartę **QUEST**.

Przycinanie drzew

W przypadku metod CRT i QUEST można uniknąć przeuczenia modelu, **przycinając** drzewo: drzewo rozrasta się do momentu spełniania kryteriów zatrzymania, a następnie jest automatycznie przycinane do najmniejszego poddrzewa na podstawie określonej maksymalnej różnicy ryzyka. Wartość ryzyka jest wyrażona w błędach standardowych. Wartością domyślną jest 1. Wartość ta musi być liczbą nieujemną. W celu uzyskania poddrzewa o minimalnej wartości ryzyka należy podać wartość 0.

W celu przycięcia drzewa

1. W głównym oknie dialogowym Drzewo decyzyjne jako metodę wzrostu wybierz **CRT** lub **QUEST**.
2. Kliknij opcję **Kryteria**.
3. Kliknij kartę **Obcinanie**.

Przycinanie a ukrywanie węzłów

Węzły usunięte z drzewa w wyniku przycięcia są niedostępne w ostatecznym drzewie. Można interaktywnie ukrywać i uwidaczniać wybrane węzły podrzędne w ostatecznym drzewie, ale nie można uwidocznić węzłów usuniętych w wyniku przycinania podczas tworzenia drzewa. Więcej informacji można znaleźć w temacie “Edytor drzewa” na stronie 16.

Substytuty

W metodach CRT i QUEST mogą być stosowane **Elementy zastępcze** dla zmiennych niezależnych (predyktora). W przypadku obserwacji, w których brak wartości dla tej zmiennej, do klasyfikacji stosowane są inne zmienne niezależne o wysokim stopniu powiązań z oryginalną zmienną. Te alternatywne predyktory są zwane elementami zastępczymi. Istnieje możliwość określenia maksymalnej liczby elementów zastępczych, jaka ma być używana w modelu.

- Domyślnie maksymalna liczba elementów zastępczych wynosi jeden minus liczba zmiennych niezależnych. Innymi słowy, dla każdej zmiennej niezależnej wszystkie pozostałe zmienne niezależne mogą być używane jako elementy zastępcze.
- Jeśli nie chcesz, aby w modelu używane były elementy zastępcze, wskaż jako liczbę elementów zastępczych wartość 0.

Aby wskazać elementy zastępcze

1. W głównym oknie dialogowym Drzewo decyzyjne jako metodę wzrostu wybierz **CRT** lub **QUEST**.
2. Kliknij opcję **Kryteria**.
3. Kliknij kartę **Elementy zastępcze**.

Opcje

To, jakie opcje są dostępne, zależeć może od metody wzrostu, poziomu pomiaru zmiennej zależnej i/lub istnienia zdefiniowanych etykiet wartości zmiennej zależnej.

Koszty błędnej klasyfikacji

W przypadku jakościowych (nominalnych, porządkowych) zmiennych zależnych koszty błędnej klasyfikacji umożliwiają uwzględnienie informacji o względnej karze za nieprawidłową klasyfikację. Na przykład:

- Koszt odmowy udzielenia kredytu klientowi, który ma zdolność kredytową, prawdopodobnie różni się od kosztu udzielenia kredytu klientowi, który później okaże się niewypłacalny.

- Koszt zaniżenia (błędnej klasyfikacji) ryzyka choroby serca u osoby, u której jest ono wysokie, będzie prawdopodobnie znacznie wyższy od kosztu zawyżenia takiego ryzyka u osoby, u której faktycznie jest ono niskie.
- Koszt wysłania przesyłki reklamowej do osoby, która prawdopodobnie na nią nie zareaguje, jest raczej niskie, natomiast koszt niewysłania jej do osoby, która prawdopodobnie zareaguje, jest względnie wyższe (w kategoriach utraconych przychodów).

Koszty błędnej klasyfikacji i etykiety wartości

Aby to okno dialogowe było dostępne, co najmniej dwie wartości jakościowej zmiennej zależnej muszą mieć zdefiniowane etykiety wartości. .

Aby określić koszty błędnej klasyfikacji

1. W głównym oknie dialogowym Drzewo klasyfikacyjne wybierz jakościową zmienną zależną (nominalną, porządkową) z co najmniej dwiema zdefiniowanymi etykietami wartości.
2. Kliknij przycisk **Opcje**.
3. Kliknij kartę **Koszty błędnej klasyfikacji**.
4. Kliknij opcję **Użytkownika**.
5. W tabeli wprowadź jeden lub większą liczbę kosztów błędnej klasyfikacji. Wartości muszą być nieujemne. (Prawidłowe klasyfikacje — wartości na przekątnej mają zawsze koszt równy 0).

Wypełnianie macierzy. W wielu przypadkach może być pożądane, aby koszty były symetryczne – czyli koszt błędnego klasyfikowania A jako B jest taki sam, jak koszt błędnego klasyfikowania B jako A. Następujące elementy sterujące mogą ułatwić określenie symetrycznej macierzy kosztów:

- **Powtórz dolny trójkąt.** Kopiuje wartości z dolnego trójkąta macierzy (pod przekątną) do odpowiednich komórek górnego trójkąta.
- **Powtórz górny trójkąt.** Kopiuje wartości z górnego trójkąta macierzy (nad przekątną) do odpowiednich komórek dolnego trójkąta.
- **Użyj średnich wartości.** Dla każdej komórki w każdej połowie macierzy dwie wartości (z górnego i dolnego trójkąta) są uśredniane, a średnia ta zastępuje każdą z tych wartości. Na przykład, jeśli koszty błędnej klasyfikacji A jako B wynosi 1, a koszt błędnej klasyfikacji B jako A wynosi 3, to użycie tej opcji spowoduje zastąpienie obu wartości średnią (1+3)/2 = 2.

Zyski

Do poziomów zależnej zmiennej jakościowej można przypisać wartości dochodów i kosztów.

- Zysk obliczany jest jako dochód minus koszt.
- Od wartości zysku zależą przedstawione w tabeli korzyści wartości średniego zysku oraz zwrotu z inwestycji. Wartości te nie mają wpływu na strukturę podstawowego modelu drzewa.
- Wartości dochodów i kosztów muszą być liczbowe i określone dla wszystkich kategorii zmiennej zależnej wyświetlanych w tabeli.

Zyski i etykiety wartości

To okno dialogowe wymaga zdefiniowanych etykiet wartości zmiennej zależnej. Aby było dostępne, co najmniej dwie wartości jakościowej zmiennej zależnej muszą mieć zdefiniowane etykiety wartości. .

Aby określić zyski

1. W głównym oknie dialogowym Drzewo klasyfikacyjne wybierz jakościową zmienną zależną (nominalną, porządkową) z co najmniej dwiema zdefiniowanymi etykietami wartości.
2. Kliknij przycisk **Opcje**.
3. Kliknij kartę **Zyski**.
4. Kliknij opcję **Użytkownika**.
5. Wprowadź dochody i koszty dla wszystkich kategorii zmiennych zależnych wymienionych w tabeli.

Prawdopodobieństwa a priori

W przypadku drzew CRT i QUEST z jakościowymi zmiennymi zależnymi można określić prawdopodobieństwa a priori przynależności do grup. **Prawdopodobieństwa a priori** to oszacowania ogólnej względnej częstości dla każdej kategorii zmiennej zależnej przed uzyskaniem jakichkolwiek informacji o wartościach zmiennej niezależnej (predyktora). Za pomocą prawdopodobieństwa a priori można skorygować wzrost drzewa spowodowany próbką, która nie jest reprezentatywna dla całej populacji.

Uzyskane z próby uczącej (empiryczne). Tego ustawienia należy użyć, jeśli rozkład wartości zmiennej zależnej w pliku danych jest reprezentatywny dla rozkładu populacji. W przypadku walidacji z podziałem próby stosowany jest rozkład obserwacji w próbie szkoleniowej.

Uwaga: Ponieważ podczas walidacji z podziałem próby obserwacje są losowo przypisywane do próby szkoleniowej, rzeczywisty rozkład obserwacji w próbie szkoleniowej nie jest z góry znany. Więcej informacji można znaleźć w temacie “Sprawdzanie” na stronie 4.

Równe dla wszystkich kategorii. Użyj tego ustawienia, jeśli kategorie zmiennej zależnej są równo reprezentowane w populacji. Na przykład, jeśli istnieją cztery kategorie, a do każdej należy około 25% obserwacji.

Użytkownika. Wprowadź nieujemną wartość dla każdej kategorii zmiennej zależnej wymienionej w tabeli. Wartości mogą być proporcjami, wartościami procentowymi, liczebnościami lub innymi wartościami reprezentującymi rozkład wartości w kategoriach.

Skoryguj aprioryczne, używając kosztów błędnej klasyfikacji. W przypadku zdefiniowania własnych kosztów błędnej klasyfikacji można skorygować prawdopodobieństwa a priori na podstawie tych kosztów. Więcej informacji można znaleźć w temacie “Koszty błędnej klasyfikacji” na stronie 8.

Zyski i etykiety wartości

To okno dialogowe wymaga zdefiniowanych etykiet wartości zmiennej zależnej. Aby było dostępne, co najmniej dwie wartości jakościowej zmiennej zależnej muszą mieć zdefiniowane etykiety wartości. .

Aby określić Prawdopodobieństwa a priori

1. W głównym oknie dialogowym Drzewo klasyfikacyjne wybierz jakościową zmienną zależną (nominalną, porządkową) z co najmniej dwiema zdefiniowanymi etykietami wartości.
2. Jako metodę wzrostu wybierz **CRT** lub **QUEST**.
3. Kliknij przycisk **Opcje**.
4. Kliknij kartę **Prawdopodobieństwa a priori**.

Oceny

W przypadku metod CHAID i Wyczerpujący CHAID z porządkową zmienną zależną można przypisywać własne oceny do poszczególnych kategorii zmiennej zależnej. Ocena oznacza rząd i odległość między kategoriami zmiennej zależnej. Za pomocą ocen można zwiększyć lub zmniejszyć względną odległość między wartościami porządkowymi lub zmienić kolejność wartości.

- **Ranga porządkowa dla każdej kategorii.** Najniższej kategorii zmiennej zależnej przypisywana jest ocena 1, drugiej co do wysokości ocena 2 itd. Jest to wartość domyślna.
- **Użytkownika.** Wprowadź wartość liczbową dla każdej kategorii zmiennej zależnej wymienionej w tabeli.

Przykład

Tabela 3. Wartości oceny określone przez użytkownika

Etykieta wartości	**Oryginalna wartość**	**Wynik**
Niewykwalifikowany	1	1
Wykwalifikowany fizyczny	2	4
Biurowy	3	4,5

Tabela 3. Wartości oceny określone przez użytkownika (kontynuacja)

Etykieta wartości	Oryginalna wartość	Wynik
Professional	4	7
Zarządzanie	5	6

- Oceny zwiększają względną odległość między kategorią *Niewykwalifikowany* a *Wykwalifikowany fizyczny* i zmniejszają względną odległość między kategorią *Wykwalifikowany fizyczny* a *Biurowy*.
- Oceny odwracają kolejność kategorii *Menedżer* i *Specjalista*.

Oceny i etykiety wartości

To okno dialogowe wymaga zdefiniowanych etykiet wartości zmiennej zależnej. Aby było dostępne, co najmniej dwie wartości jakościowej zmiennej zależnej muszą mieć zdefiniowane etykiety wartości. .

Aby określić oceny

1. W głównym oknie dialogowym Drzewo klasyfikacyjne wybierz porządkową zmienną zależną z co najmniej dwiema zdefiniowanymi etykietami wartości.
2. Jako metodę wzrostu wybierz **CHAID** lub **Wyczerpujący CHAID**.
3. Kliknij przycisk **Opcje**.
4. Kliknij kartę **Oceny**.

Braki danych

Karta Braki danych zawiera ustawienia wpływające na sposób traktowania braków danych użytkownika w nominalnych zmiennych niezależnych (predyktorach).

- Sposób traktowania braków danych użytkownika w porządkowych i ilościowych zmiennych niezależnych zależy od metody wzrostu.
- Sposób postępowania z nominalnymi zmiennymi zależnymi określa się w oknie dialogowym Kategorie. Więcej informacji można znaleźć w temacie “Wybór kategorii” na stronie 4.
- W przypadku porządkowych i ilościowych zmiennych zależnych obserwacje z systemowymi brakami danych i brakami danych użytkownika zawsze są wykluczane.

Traktuj jako brakujące wartości. Braki danych użytkownika są traktowane jak systemowe braki danych. Postępowanie z systemowymi brakami danych zależy od metody wzrostu.

Traktuj jako ważne wartości. Braki danych użytkownika w nominalnej zmiennej niezależnej są traktowane jako ważne obserwacje podczas budowania drzewa i klasyfikacji obserwacji.

Reguły zależne od metody

Jeśli niektóre, lecz nie wszystkie wartości zmiennej zależnej są systemowymi brakami danych lub brakami danych użytkownika:

- W metodach CHAID i Wyczerpujący CHAID systemowe braki danych oraz braku danych użytkownika w zmiennych niezależnych są uwzględniane w analizie jako jedna kategoria łączna. W przypadku ilościowych i porządkowych zmiennych zależnych algorytmy najpierw generują kategorie na podstawie ważnych wartości, a następnie decydują, czy scalić kategorię braków z najbardziej podobną kategorią ważnych wartości, czy też zachować ją jako oddzielną kategorię.
- W metodach CRT i QUEST obserwacje z brakami danych zmiennej niezależnej są wykluczane z procesu wzrostu drzewa, ale są klasyfikowane z pomocą substytutów, jeśli substytuty są uwzględnione w metodzie. Jeśli braki danych użytkownika w zmiennej nominalnej są traktowane jako brakujące, to również są w ten sposób traktowane. Więcej informacji można znaleźć w temacie “Substytuty” na stronie 8.

Aby określić sposób traktowania braków danych użytkownika w niezależnych zmiennych nominalnych

1. W głównym oknie dialogowym Drzewo klasyfikacyjne wybierz co najmniej jedną nominalną zmienną niezależną.
2. Kliknij przycisk **Opcje**.
3. Kliknij kartę **Braki danych**.

Zapisywanie informacji o modelu

Istnieje możliwość zapisywania informacji z modelu jako zmiennych w roboczym pliku danych, a także zapisanie całego modelu w formacie XML (PMML) w pliku zewnętrznym.

Zapisywane zmienne

Numer węzła końcowego. Węzeł końcowy, do którego przypisana jest każda z obserwacji. Wartość jest numerem węzła drzewa.

Wartość przewidywana. Klasa (grupa) lub wartość zmiennej zależnej przewidywana przez model.

Przewidywane prawdopodobieństwa. Prawdopodobieństwo powiązane z predykcją modelu. Dla każdej kategorii zmiennej zależnej zapisywana jest jedna zmienna. Opcja niedostępna dla ilościowych zmiennych zależnych.

Przypisanie do próby (ucząca/testująca). W przypadku walidacji z podziałem próby ta zmienna określa, czy dana obserwacja była używana w próbie uczącej, czy testującej. Wartość 1 oznacza próbę uczącą, a 0 oznacza próbę testującą. Opcja niedostępna, jeśli nie wybrano walidacji z podziałem próby. Więcej informacji można znaleźć w temacie “Sprawdzanie” na stronie 4.

Eksportowanie modelu drzewa jako pliku XML

Istnieje możliwość zapisania całego modelu drzewa w formacie XML (PMML). Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. .

Próba ucząca. Zapisuje model w określonym pliku. W przypadku drzew z walidacją z podziałem próby jest to model dla próby uczącej.

Próba testująca. Zapisuje model próby testującej w określonym pliku. Opcja niedostępna, jeśli nie wybrano walidacji z podziałem próby.

Dane wyjściowe

Dostępne opcje raportów zależą od metody wzrostu, poziomu pomiaru zmiennej niezależnej oraz innych ustawień.

Widok drzewa

Można wpłynąć na początkowy wygląd drzewa lub całkowicie wyłączyć wyświetlanie drzewa.

Drzewo. Domyślnie diagram drzewa jest uwzględniony w wynikach prezentowanych w oknie raportu. Usuń zaznaczenie tej opcji, aby diagram drzewa nie był uwzględniany w wynikach.

Pokaż. Te opcje sterują początkowym wyglądem diagramu drzewa w oknie raportu. Wszystkie te atrybuty można także modyfikować, edytując wygenerowane drzewo.

- **Orientacja.** Drzewo może być wyświetlane zstępująco, z węzłem głównym u góry, od lewej do prawej albo od prawej do lewej.
- **Zawartość węzłów w postaci.** Węzły mogą być wyświetlane jako tabele, wykresy lub w obu tych postaciach jednocześnie. W przypadku jakościowych zmiennych zależnych tabele zawierają liczebności i wartości procentowe, a wykresy są wykresami słupkowymi. W przypadku ilościowych zmiennych zależnych tabele zawierają średnie, odchylenia standardowe, liczby obserwacji i wartości przewidywane, a wykresy są histogramami.
- **Powiększenie.** Domyślnie duże drzewa są automatycznie zmniejszane tak, by mieściły się na stronie. Można określić własną skalę w procentach, maksymalnie 200%.

- **Statystyki zmiennej niezależnej.** W przypadku metod CHAID i Wyczerpujący CHAID statystyki obejmują wartość *F* (dla ilościowych zmiennych zależnych) lub chi-kwadrat (dla jakościowych zmiennych zależnych), a także wartość istotności i liczbę stopni swobody. W przypadku metody CRT podawana jest wartość ulepszenia. W przypadku metody QUEST podawana jest wartość *F*, wartość istotności i liczba stopni swobody dla ilościowych i porządkowych zmiennych niezależnych; dla nominalnych zmiennych niezależnych podawana jest wartość chi-kwadrat, wartość istotności i liczba stopni swobody.
- **Definicje węzłów.** W definicjach węzłów podawane są wartości zmiennej niezależnej używane przy każdym podziale węzła.

Drzewo w postaci tabeli. Informacje podsumowujące dotyczące każdego węzła w drzewie, w tym numer węzła nadrzędnego, statystyki zmiennej zależnej, wartości zmiennych niezależnych dla węzła, średnia i odchylenie standardowe ilościowych zmiennych zależnych lub liczebności i wartości procentowe jakościowych zmiennych zależnych.

Aby określić początkowy sposób wyświetlania drzewa

1. W głównym oknie dialogowym Drzewo decyzyjne kliknij opcję **Wynik**.
2. Kliknij kartę **Drzewo**.

Statystyki

To, które tabele statystyk są dostępne, zależy od poziomu pomiaru zmiennej zależnej, metody wzrostu i innych ustawień.

Model

Podsumowanie. Podsumowanie obejmuje informacje o zastosowanej metodzie, zmiennych uwzględnionych w modelu oraz zmiennych określonych, ale nieuwzględnionych w modelu.

Ryzyko. Oszacowanie ryzyka i jego błąd standardowy. Miara dokładności predykcji drzewa.

- W przypadku jakościowych zmiennych zależnych oszacowanie ryzyka jest odsetkiem nieprawidłowo sklasyfikowanych obserwacji po korekcie z uwzględnieniem prawdopodobieństwa a priori i kosztów błędnej klasyfikacji.
- W przypadku ilościowych zmiennych zależnych oszacowanie ryzyka jest równe wariancji wewnątrz węzła.

Tabela klasyfikacji. W przypadku jakościowych (nominalnych, porządkowych) zmiennych zależnych tabela ta zawiera liczby obserwacji sklasyfikowanych prawidłowo i nieprawidłowo dla każdej kategorii zmiennej zależnej. Opcja niedostępna dla ilościowych zmiennych zależnych.

Koszt, prawdopodobieństwo wstępne, ocena i zysk. W przypadku jakościowych zmiennych zależnych tabela ta przedstawia koszty, prawdopodobieństwa a priori, oceny i zyski używane w analizie. Opcja niedostępna dla ilościowych zmiennych zależnych.

Zmienne niezależne

Ważność predyktora dla modelu. W przypadku metody wzrostu CRT szereguje zmienne niezależne (predyktory) według ich ważności dla modelu. Opcja niedostępna w przypadku metod QUEST i CHAID.

Substytucyjne według podziału. W przypadku metod wzrostu CRT i QUEST, jeśli model zawiera substytuty, prezentuje listę substytutów dla każdego podziału w drzewie. Opcja niedostępna w przypadku metody CHAID. Więcej informacji zawiera temat “Substytuty” na stronie 8.

Wydajność węzłów

Podsumowanie. W przypadku ilościowych zmiennych zależnych tabela zawiera numer węzła, liczbę obserwacji i średnią zmiennej zależnej. W przypadku jakościowych zmiennych zależnych ze zdefiniowanymi zyskami tabela zawiera numer węzła, liczbę obserwacji, średni zysk i zwrot z inwestycji. Opcja niedostępna dla jakościowych zmiennych zależnych bez zdefiniowanych zysków. Więcej informacji zawiera temat “Zyski” na stronie 9.

Według kategorii docelowej. W przypadku jakościowych zmiennych zależnych ze zdefiniowanymi kategoriami docelowymi tabela zawiera procentową korzyść, wartość procentową odpowiedzi i wartość procentową indeksu (przyrost) z podziałem na węzły lub grupy percentylowe. Dla każdej kategorii docelowej generowana jest odrębna tabela. Opcja niedostępna w przypadku ilościowych zmiennych zależnych lub jakościowych zmiennych zależnych bez zdefiniowanych kategorii docelowych. Więcej informacji zawiera temat "Wybór kategorii" na stronie 4.

Wiersze. Tabele wydajności węzłów mogą przedstawiać wyniki według węzłów końcowych i/lub percentyli. W wypadku wybrania obu tych sposobów prezentacji dla każdej kategorii docelowej generowane są dwie tabele. W tabelach percentyli prezentowane są wartości skumulowane dla każdego percentyla, na podstawie kolejności sortowania.

Przyrost percentylowy. Dla tabel percentyli można wybrać przyrost: 1, 2, 5, 10, 20 lub 25.

Statystyki wartości skumulowanych. Powoduje, że tabele węzłów końcowych zawierają dodatkowe kolumny z wynikami skumulowanymi.

Aby wybrać statystyki wynikowe

1. W głównym oknie dialogowym Drzewo decyzyjne kliknij opcję **Wynik**.
2. Kliknij kartę **Statystyki**.

Wykresy

To, które wykresy są dostępne, zależy od poziomu pomiaru zmiennej zależnej, metody wzrostu i innych ustawień.

Ważność zmiennej niezależnej dla modelu. Wykres słupkowy ważności modelu według zmiennych niezależnych (predyktorów). Opcja dostępna tylko w przypadku metody wzrostu CRT.

Wydajność węzłów

Korzyść. Korzyść to odsetek łącznej liczby obserwacji w kategorii docelowej w każdym węźle, obliczany wg wzoru: (*n* w docelowej węzła / razem *n* w docelowej) x 100. Wykres korzyści jest liniowym wykresem skumulowanych korzyści dla percentyli, obliczanym wg wzoru: (skumulowane *n* w docelowej dla percentyla / razem *n* w docelowej) x 100. Dla każdej kategorii docelowej generowany jest odrębny wykres liniowy. Opcja dostępna tylko w przypadku jakościowych zmiennych zależnych ze zdefiniowanymi kategoriami docelowymi. Więcej informacji można znaleźć w temacie "Wybór kategorii" na stronie 4.

Wykres korzyści przedstawia wartości, które byłyby zawarte w kolumnie procentowej wartości korzyści w tabeli percentyli, która zawiera również wartości skumulowane.

Indeks. Indeks stanowi proporcję procentową odpowiedzi węzła dla kategorii wynikowej, porównaną do całkowitej wartości procentowej odpowiedzi kategorii dla całej próbki. Wykres indeksu jest liniowym wykresem skumulowanych wartości indeksu percentyla. Dostępny jest tylko w przypadku jakościowych zmiennych zależnych. Skumulowany indeks dla percentyla obliczany jest ze wzoru: (skumulowana wartość procentowa odpowiedzi dla percentyla / łączna wartość procentowa odpowiedzi) x 100. Dla każdej kategorii docelowej generowany jest odrębny wykres, a kategorie docelowe muszą być zdefiniowane.

Wykres indeksu przedstawia wartości, które byłyby zawarte w kolumnie *Indeks* w tabeli korzyści dla percentyli.

Odpowiedź. Procent obserwacji w węźle w określonej kategorii wynikowej. Wykres odpowiedzi jest liniowym wykresem skumulowanego percentyla odpowiedzi, obliczonego jako: (cel skumulowanego percentyla *n* / wartość skumulowanego percentyla łącznie *n*) x 100. Opcja dostępna tylko w przypadku jakościowych zmiennych zależnych ze zdefiniowanymi kategoriami docelowymi.

Wykres odpowiedzi przedstawia wartości, które byłyby zawarte w kolumnie *Odpowiedź* w tabeli korzyści dla percentyli.

Średnia. Wykres liniowy skumulowanych średnich dla percentyli dla zmiennej zależnej. Dostępny jest tylko w przypadku ilościowych zmiennych zależnych.

Przeciętny zysk. Wykres liniowy skumulowanego średniego zysku. Wykres dostępny tylko w przypadku jakościowych zmiennych zależnych ze zdefiniowanymi zyskami. Więcej informacji zawiera temat “Zyski” na stronie 9.

Wykres przeciętnego zysku przedstawia wartości, które byłyby zawarte w kolumnie *Zysk* w tabeli podsumowania korzyści dla percentyli.

Zwrot z inwestycji (ROI). Wykres liniowy skumulowanego zwrotu z inwestycji (ROI — return on investment). Zwrot z inwestycji oblicza się jako stosunek zysków do kosztów. Wykres dostępny tylko w przypadku jakościowych zmiennych zależnych ze zdefiniowanymi zyskami.

Wykres zwrotu z inwestycji przedstawia wartości, które byłyby zawarte w kolumnie *ROI* w tabeli podsumowania korzyści dla percentyli.

Przyrost percentylowy. W odniesieniu do wszystkich wykresów percentylowych określa przyrost uwzględniony na wykresie: 1, 2, 5, 10, 20 albo 25.

Aby wybrać wykres wynikowy

1. W głównym oknie dialogowym Drzewo decyzyjne kliknij opcję **Wynik**.
2. Kliknij kartę **Wykresy**.

Reguły wyboru i oceniania

Karta Reguły umożliwia generowanie reguł wyboru lub klasyfikacji/predykcji w formie komend, instrukcji SQL lub zwykłego tekstu (w języku angielskim). Reguły te można wyświetlać w oknie raportu i/lub zapisać w pliku zewnętrznym.

Składnia. Określa postać reguł wyboru w obu zbiorach wyników: wyświetlanych w oknie raportów oraz zapisanych w pliku zewnętrznym.

- **IBM SPSS Statistics.** Język składni komend. Reguły są wyrażone jako zbiór komend definiujących warunek filtru, który można wykorzystać do wybierania podzbiorów obserwacji lub jako instrukcję COMPUTE, która może być używana do oceniania obserwacji.
- **SQL.** Standardowe reguły SQL generowane są w celu wybierania lub wyodrębniania z bazy danych rekordów lub przypisywania wartości do tych rekordów. Wygenerowane reguły SQL nie obejmują żadnych nazw tabel ani innych źródeł danych.
- **Prosty tekst.** Pseudokod w języku angielskim. Reguły wyrażone są w formie szeregu instrukcji logicznych „if...then”, które opisują sposób, w jaki model realizuje klasyfikację lub predykcję każdego węzła. Reguły w tej postaci mogą zawierać zdefiniowane etykiety zmiennych i wartości lub nazwy zmiennych i wartości danych.

Typ. W przypadku reguł IBM SPSS Statistics i SQL określa typ generowanych reguł: reguły wyboru albo reguły oceniania.

- **Przypisywanie wartości do obserwacji.** Reguły mogą być używane do przypisywania predykcji modelu do obserwacji spełniających kryteria członkostwa w węźle. Dla każdego węzła generowana jest osobna reguła spełniająca kryteria członkostwa w węźle.
- **Wybieranie obserwacji.** Reguły można wykorzystać do wybierania obserwacji spełniających kryteria członkostwa w węźle. W przypadku IBM SPSS Statistics i reguł SQL generowana jest pojedyncza reguła umożliwiająca wybór wszystkich obserwacji spełniających kryteria wyboru.

Dołącz predyktory substytucyjne w regułach IBM SPSS Statistics i SQL. W przypadku CRT i QUEST istnieje możliwość uwzględniania w regułach zastępczych predyktorów z modelu. Reguły uwzględniania elementów zastępczych mogą być dość złożone. W ogólnym wypadku w przypadku chęci czerpania informacji koncepcyjnych dotyczących drzewa, elementy zastępcze należy wykluczyć. Jeśli niektóre obserwacje mają niekompletne dane zmiennej niezależnej (predyktora), a użytkownik chciałby, aby reguły odzwierciedlały jego drzewo, powinien uwzględnić elementy zastępcze. Więcej informacji zawiera temat “Substytuty” na stronie 8.

Węzły. Steruje zasięgiem generowanych reguł. Dla każdego węzła w zasięgu generowana jest osobna reguła.

- **Wszystkie węzły końcowe.** Generuje reguły dla wszystkich węzłów końcowych.
- **Najlepsze węzły końcowe.** Generuje reguły dla *n* najlepszych węzłów końcowych (na podstawie wartości indeksów). Jeśli liczba przekracza liczbę węzłów końcowych w drzewie, reguły są generowane dla wszystkich węzłów końcowych. (Patrz uwaga poniżej).
- **Najlepsze węzły końcowe do podanego procentu obserwacji.** Generuje reguły dla węzłów końcowych obejmujących *n* procent najlepszych obserwacji (na podstawie wartości indeksów). (Patrz uwaga poniżej).
- **Węzły końcowe z indeksem równym lub większym od wartości granicznej.** Generuje reguły dla wszystkich węzłów końcowych z wartością indeksu większą lub równą podanej. Wartość indeksu większa od 100 oznacza, że odsetek obserwacji w kategorii docelowej tego węzła przekracza odsetek w węźle głównym. (Patrz uwaga poniżej).
- **Wszystkie węzły.** Generuje reguły dla wszystkich węzłów.

Uwaga 1: Wybór węzłów na podstawie wartości indeksów jest dostępny tylko w przypadku jakościowych zmiennych zależnych ze zdefiniowanymi kategoriami docelowymi. Jeśli określono więcej niż jedną kategorię docelową, to dla każdej z nich generowany jest odrębny zestaw reguł.

Note 2: For IBM SPSS Statistics and SQL rules for selecting cases (not rules for assigning values), **All nodes** and **All terminal nodes** will effectively generate a rule that selects all cases used in the analysis.

Eksportuj reguły do pliku. Zapisuje reguły w zewnętrznym pliku tekstowym.

Można też generować i zapisywać reguły wyboru lub oceniania interaktywnie, na podstawie wyboru węzłów w ostatecznym modelu drzewa. Więcej informacji można znaleźć w temacie “Reguły wyboru i oceniania obserwacji” na stronie 19.

Uwaga: W przypadku zastosowania reguł w postaci składni komendy do innego pliku danych ten plik danych musi zawierać zmienne o tej samej nazwie, co zmienne niezależne uwzględnione w modelu końcowym, zmierzone za pomocą tej samej metryki, o tych samych brakach danych zdefiniowanych przez użytkownika (o ile występują).

Aby określić reguły wyboru lub oceniania

1. W głównym oknie dialogowym Drzewo decyzyjne kliknij opcję **Wynik**.
2. Kliknij kartę **Reguły**.

Edytor drzewa

Korzystając z Edytora drzewa, można:

- Ukrywać i wyświetlać wybrane gałęzie drzewa.
- Sterować wyświetlaniem zawartości węzła, statystyk wyświetlanych przy podziale węzła i innych informacji.
- Zmieniać kolor węzłów, tła, krawędzi, wykresów i czcionek.
- Zmieniać styl i rozmiar czcionki.
- Zmieniać wyrównanie drzewa.
- Wybierać podzbiory obserwacji do dalszej analizy w wybranych węzłach.
- Tworzyć i zapisywać reguły wyboru lub oceniania obserwacji w oparciu o wybrane węzły.

Aby edytować model drzewa:

1. Dwukrotnie kliknij model drzewa w oknie przeglądarki.

   lub
2. W menu Edycja lub w menu podręcznym dostępnym po kliknięciu prawym przyciskiem myszy wybierz:

   Edytuj

Ukrywanie i wyświetlanie węzłów

Aby ukryć (zwinąć) wszystkie elementy podrzędne w rozgałęzieniu poniżej węzła nadrzędnego:

1. Kliknij symbol minus (–) w niewielkim prostokącie poniżej dolnego prawego rogu węzła nadrzędnego.

   Wszystkie węzły poniżej węzła nadrzędnego w tym rozgałęzieniu zostaną ukryte.

   Aby wyświetlić (rozwinąć) wszystkie elementy podrzędne w rozgałęzieniu poniżej węzła nadrzędnego:
2. Kliknij symbol plus (+) w niewielkim prostokącie poniżej dolnego prawego rogu węzła nadrzędnego.

Uwaga: Ukrycie węzłów podrzędnych w rozgałęzieniu nie jest równoważne obcięciu drzewa. W celu uzyskania drzewa obciętego należy zaznaczyć to jeszcze przed jego utworzeniem. Obcięte rozgałęzienia nie będą wówczas uwzględniane w końcowym modelu drzewa. Więcej informacji można znaleźć w temacie "Przycinanie drzew" na stronie 8.

Wybieranie wielu węzłów

W oparciu o obecnie wybrane węzły można wybierać obserwacje, generować reguły oceniania i wyboru oraz wykonywać inne czynności. Aby wybrać wiele węzłów:

1. Kliknij węzeł, który chcesz wybrać.
2. Kliknij z wciśniętym klawiszem Ctrl inną właściwość, którą chcesz scalić.

Można wybrać wiele węzłów równorzędnych i/lub węzły nadrzędne w jednym rozgałęzieniu oraz węzły podrzędne w innym rozgałęzieniu. Nie można jednak zastosować wielokrotnego wyboru do węzła nadrzędnego oraz podrzędnego/potomnego w tym samym rozgałęzieniu węzła.

Praca z dużymi drzewami

Modele drzewa mogą niekiedy zawierać tak wiele węzłów i rozgałęzień, że wyświetlenie całego drzewa w pełnej wielkości staje się trudne lub niemożliwe. Istnieje szereg funkcji, które mogą okazać się użyteczne podczas pracy z dużymi drzewami:

- **Mapa drzewa.** Mapa drzewa, znacznie mniejsza i prostsza jego wersja, umożliwia nawigowanie po drzewie i wybór poszczególnych węzłów. Więcej informacji można znaleźć w temacie "Mapa drzewa" na stronie 17.
- **Skalowanie.** Umożliwia powiększanie i pomniejszanie poprzez zmianę wartości procentowej skali dla ekranu drzewa. Więcej informacji można znaleźć w temacie "Skalowanie ekranu drzewa" na stronie 18.
- **Ekran węzła i rozgałęzienia.** Drzewo można uczynić bardziej kompaktowym, wyświetlając w węzłach tylko tabele lub tylko wykresy i/lub ukrywając wyświetlanie etykiet węzłów lub informacji o zmiennych niezależnych. Więcej informacji można znaleźć w temacie "Sterowanie informacjami wyświetlanymi w drzewie" na stronie 18.

Mapa drzewa

Mapa drzewa oferuje kompaktowy, uproszczony widok drzewa, którego można użyć do nawigacji w drzewie i wyboru węzłów.

Aby użyć okna mapy drzewa:

1. Z menu Edytora drzewa wybierz kolejno następujące pozycje:

   Widok > **Mapa drzewa**

- Obecnie wybrany węzeł jest podświetlony zarówno w Edytorze modelu drzewa, jak i w oknie mapy drzewa.
- Część drzewa widoczna obecnie w obszarze widoku Edytora modelu drzewa jest na mapie drzewa oznaczona czerwonym prostokątem. Kliknij prawym przyciskiem myszy, a następnie przeciągnij prostokąt, aby zmienić sekcję drzewa wyświetlaną w obszarze widoku.
- Po wybraniu węzła na mapie drzewa, które nie znajduje się obecnie w obszarze widoku Edytora drzewa widoki ulegają przesunięciu tak, aby uwzględniały wybrany węzeł.

- Wybór wielu węzłów działa tak samo na mapie drzewa, jak w Edytorze drzew: kliknięcie przy wciśniętym klawiszu Ctrl umożliwia wybór wielu węzłów. Nie można zastosować wielokrotnego wyboru do węzła nadrzędnego oraz podrzędnego/potomnego w tym samym rozgałęzieniu węzła.

Skalowanie ekranu drzewa

Domyślnie drzewa są automatycznie skalowane w celu dopasowania do okna raportów, co może skutkować początkowymi trudnościami w odczycie niektórych drzew. Istnieje możliwość wyboru wstępnie ustawionego ustawienia skali lub wprowadzenia własnej niestandardowej wartości skali z przedziału od 5% do 200%.

Aby zmienić skalę drzewa:

1. Wybierz wartość procentową skali z rozwijanej listy na pasku narzędzi lub wprowadź niestandardową wartość procentową.

   lub

2. Z menu Edytora drzewa wybierz kolejno następujące pozycje:

   Widok > **Skala...**

Można także wskazać wartość skali przed utworzeniem modelu drzewa. Więcej informacji można znaleźć w temacie “Dane wyjściowe” na stronie 12.

Okno Podsumowanie węzła

Okno podsumowania węzła udostępnia większy widok wybranych węzłów. Okna podsumowania można także użyć w celu wyświetlenia, zastosowania lub zapisania reguł wyboru lub oceniania w oparciu o wybrane węzły.

- Użyj menu Widok w oknie podsumowania węzła w celu przełączania się między widokami tabeli podsumowań, tabeli przestawnej oraz reguł.
- Okno Reguły w oknie podsumowania węzła umożliwia wybór typu reguł, które mają zostać wyświetlone. Więcej informacji można znaleźć w temacie “Reguły wyboru i oceniania obserwacji” na stronie 19.
- Wszystkie widoki w oknie podsumowania węzła odzwierciedlają połączone podsumowanie dla wszystkich wybranych węzłów.

W celu użycia okna podsumowania węzła:

1. Wybierz węzły w Edytorze drzew. Aby wybrać wiele węzłów, należy je kliknąć z naciśniętym klawiszem Ctrl.
2. Z menu wybierz:

   Widok > **Podsumowanie**

Sterowanie informacjami wyświetlanymi w drzewie

Menu Opcje w Edytorze drzewa umożliwia sterowanie wyświetlaniem zawartości węzła, nazwami i statystykami zmiennych niezależnych (predyktora), definicjami węzłów oraz innymi ustawieniami. Wieloma z tych ustawień można również sterować z paska narzędzi.

Zmiana kolorów drzewa i czcionek tekstu

Istnieje możliwość zmiany następujących kolorów w drzewie:

- Krawędź węzła, tła i kolor tekstu
- Kolor rozgałęzienia i kolor tekstu rozgałęzienia
- Kolor tła drzewa
- Predykowany kolor podświetlenia kategorii (zmienne zależne kategorialne)
- Kolory wykresu węzłów

Można także zmienić czcionkę, styl i wielkość wszystkich tekstów w drzewie.

Uwaga: Nie da się zmieniać kolorów ani atrybutów czcionek poszczególnych węzłów ani gałęzi. Zmiany kolorów zostają zastosowane do wszystkich elementów tego samego typu, a zmiany czcionek (inne niż kolor) dotyczą wszystkich elementów wykresu.

Aby zmienić atrybuty i kolory czcionki tekstu:

1. Użyj paska narzędzi w celu zmiany atrybutów czcionki w całym drzewie lub kolorów poszczególnych elementów drzewa. (Dla każdego wskazanego kursorem myszy elementu sterującego na pasku narzędzi wyświetlana jest opisująca go etykietka).

   lub

2. Dwukrotnie kliknij w dowolnym miejscu w Edytorze drzew, aby otworzyć okno Właściwości:

   Widok > **Właściwości**

3. Aby zmienić atrybuty krawędzi, gałęzi, tła węzła, przewidywanej kategorii i tła drzewa, kliknij kartę **Kolor**.
4. Aby uzyskać dostęp do kolorów i atrybutów, kliknij opcję **Tekst**.
5. Aby uzyskać dostęp do kolorów wykresu węzłów, kliknij kartę **Wykresy węzłów**.

Reguły wyboru i oceniania obserwacji

Edytor drzewa umożliwia:

- Wybierz podzbiory obserwacji w oparciu o wybrane węzły. Więcej informacji można znaleźć w temacie “Filtrowanie obserwacji” na stronie 19.
- Wygeneruj reguły wyboru obserwacji lub reguły oceniania w składni komendy IBM SPSS Statistics lub w formacie SQL. Więcej informacji można znaleźć w temacie “Zapisywanie dokonanych wyborów i reguł oceniania” na stronie 19.

Możesz również automatycznie zapisywać reguły w oparciu o różne kryteria podczas uruchamiania procedury Drzewo decyzyjne w celu utworzenia modelu drzewa. Więcej informacji można znaleźć w temacie “Reguły wyboru i oceniania” na stronie 15.

Filtrowanie obserwacji

Aby dowiedzieć się więcej o obserwacjach w danym węźle lub w danej grupie węzłów, można wybrać podzbiór rekordów do dalszej analizy na podstawie wybranych węzłów.

1. Wybierz węzły w Edytorze drzew. Aby wybrać wiele węzłów, należy je kliknąć z naciśniętym klawiszem Ctrl.
2. Wybierz z menu następującą opcję:

   Reguły > **Filtruj obserwacje...**

3. Wprowadź nazwę zmiennej filtrującej. W przypadku obserwacji z wybranych węzłów zmienna ta przyjmie wartość 1. Dla wszystkich innych obserwacji zmienna ta będzie miała wartość 0, zostanie więc wyłączona z dalszej analizy, chyba że zostanie zmieniony status filtra.
4. Kliknij **OK**.

Zapisywanie dokonanych wyborów i reguł oceniania

Istnieje możliwość zapisu obserwacji oraz reguł oceniania w pliku zewnętrznym, a następnie zastosowania ich do różnych źródeł danych. Reguły bazują na wybranych węzłach w Edytorze drzew.

Składnia. Określa postać reguł wyboru w obu zbiorach wyników: wyświetlanych w oknie raportów oraz zapisanych w pliku zewnętrznym.

- **IBM SPSS Statistics.** Język składni komend. Reguły są wyrażone jako zbiór komend definiujących warunek filtru, który można wykorzystać do wybierania podzbiorów obserwacji lub jako instrukcję `COMPUTE`, która może być używana do oceniania obserwacji.

- **SQL.** Standardowe reguły SQL generowane są w celu wybierania/wyodrębniania z bazy danych rekordów lub przypisywania wartości do tych rekordów. Wygenerowane reguły SQL nie obejmują żadnych nazw tabel ani innych źródeł danych.

Typ. Istnieje możliwość tworzenia reguł wyboru i oceniania.

- **Wybieranie obserwacji.** Reguły można wykorzystać do wybierania obserwacji spełniających kryteria członkostwa w węźle. W przypadku IBM SPSS Statistics i reguł SQL generowana jest pojedyncza reguła umożliwiająca wybór wszystkich obserwacji spełniających kryteria wyboru.
- **Przypisywanie wartości do obserwacji.** Reguły mogą być używane do przypisywania predykcji modelu do obserwacji spełniających kryteria członkostwa w węźle. Dla każdego węzła generowana jest osobna reguła spełniająca kryteria członkostwa w węźle.

Uwzględnianie elementów zastępczych. W przypadku CRT i QUEST istnieje możliwość uwzględniania w regułach zastępczych predyktorów z modelu. Reguły uwzględniania elementów zastępczych mogą być dość złożone. W ogólnym wypadku w przypadku chęci czerpania informacji koncepcyjnych dotyczących drzewa, elementy zastępcze należy wykluczyć. Jeśli niektóre obserwacje mają niekompletne dane zmiennej niezależnej (predyktora), a użytkownik chciałby, aby reguły odzwierciedlały jego drzewo, powinien uwzględnić elementy zastępcze. Więcej informacji zawiera temat “Substytuty” na stronie 8.

W celu zapisania reguł wyboru obserwacji lub oceniania:

1. Wybierz węzły w Edytorze drzew. Aby wybrać wiele węzłów, należy je kliknąć z naciśniętym klawiszem Ctrl.
2. Z menu wybierz:

   Reguły > **Eksportuj...**
3. Wybierz żądany typ reguł i wprowadź nazwę pliku.

Uwaga: W przypadku zastosowania reguł w postaci składni komendy do innego pliku danych ten plik danych musi zawierać zmienne o tej samej nazwie, co zmienne niezależne uwzględnione w modelu końcowym, zmierzone za pomocą tej samej metryki, o tych samych brakach danych zdefiniowanych przez użytkownika (o ile występują).

Uwagi

Niniejsza publikacja została przygotowana z myślą o produktach i usługach oferowanych w Stanach Zjednoczonych. IBM może udostępniać ten materiał w innych językach. Jednakże w celu uzyskania dostępu do takiego materiału istnieje konieczność posiadania egzemplarza produktu w takim języku.

Produktów, usług lub opcji opisywanych w tym dokumencie IBM nie musi oferować we wszystkich krajach. Informacje o produktach i usługach dostępnych w danym kraju można uzyskać od lokalnego przedstawiciela IBM. Odwołanie do produktu, programu lub usługi IBM nie oznacza, że można użyć wyłącznie tego produktu, programu lub usługi IBM. Zamiast nich można zastosować ich odpowiednik funkcjonalny pod warunkiem że nie narusza to praw własności intelektualnej IBM. Jednakże cała odpowiedzialność za ocenę przydatności i sprawdzenie działania produktu, programu lub usługi pochodzących od producenta innego niż IBM spoczywa na użytkowniku.

IBM może posiadać patenty lub złożone wnioski patentowe na produkty, o których mowa w niniejszej publikacji. Przedstawienie tej publikacji nie daje żadnych uprawnień licencyjnych do tychże patentów. Pisemne zapytania w sprawie licencji można przesyłać na adres:

IBM Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119
Armonk, NY 10504-1785U.S.A.

Zapytania dotyczące zestawów znaków dwubajtowych (DBCS) należy kierować do lokalnych działów własności intelektualnej IBM (IBM Intellectual Property Department) lub wysłać je na piśmie na adres:

Intellectual Property Licensing

Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokio 103-8510, Japonia

INTERNATIONAL BUSINESS MACHINES CORPORATION DOSTARCZA TĘ PUBLIKACJĘ W STANIE, W JAKIM SIĘ ZNAJDUJE ("AS IS") BEZ UDZIELANIA JAKICHKOLWIEK GWARANCJI (W TYM TAKŻE RĘKOJMI), WYRAŹNYCH LUB DOMNIEMANYCH, A W SZCZEGÓLNOŚCI DOMNIEMANYCH GWARANCJI PRZYDATNOŚCI HANDLOWEJ, PRZYDATNOŚCI DO OKREŚLONEGO CELU ORAZ GWARANCJI, ŻE PUBLIKACJA NIE NARUSZA PRAW STRON TRZECICH. Ustawodawstwa niektórych krajów nie dopuszczają zastrzeżeń dotyczących gwarancji wyraźnych lub domniemanych w odniesieniu do pewnych transakcji; w takiej sytuacji powyższe zdanie nie ma zastosowania.

Informacje zawarte w tej publikacji mogą zawierać nieścisłości techniczne lub błędy drukarskie. Informacje te są okresowo aktualizowane, a zmiany te zostaną uwzględnione w kolejnych wydaniach tej publikacji. IBM zastrzega sobie prawo do wprowadzania ulepszeń i/lub zmian w produktach i/lub programach opisanych w tej publikacji w dowolnym czasie, bez wcześniejszego powiadomienia.

Wszelkie wzmianki w tej publikacji na temat stron internetowych firm innych niż IBM zostały wprowadzone wyłącznie dla wygody użytkowników i w żadnym razie nie stanowią zachęty do ich odwiedzania. Materiały dostępne na tych stronach nie są częścią materiałów opracowanych dla tego produktu IBM, a użytkownik korzysta z nich na własną odpowiedzialność.

IBM ma prawo do używania i rozpowszechniania informacji przysłanych przez użytkownika w dowolny sposób, jaki uzna za właściwy, bez żadnych zobowiązań wobec ich autora.

Licencjobiorcy tego programu, którzy chcieliby uzyskać informacje na temat programu w celu: (i) umożliwienia wymiany informacji między niezależnie utworzonymi programami i innymi programami

(łącznie z opisywanym) oraz (ii) wykorzystywania wymienianych informacji, powinni skontaktować się z:

IBM Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119
Armonk, NY 10504-1785U.S.A.

Informacje takie mogą być udostępnione, o ile spełnione zostaną odpowiednie warunki, w tym, w niektórych przypadkach, zostanie uiszczona stosowna opłata.

Licencjonowany program opisany w niniejszej publikacji oraz wszystkie inne licencjonowane materiały dostępne dla tego programu są dostarczane przez IBM na warunkach określonych w Umowie IBM z Klientem, Międzynarodowej Umowie Licencyjnej IBM na Program lub w innych podobnych umowach zawartych między IBM i użytkownikami.

Dane dotyczące wydajności i cytowane przykłady zostały przedstawione jedynie w celu zobrazowania sytuacji. Faktyczne wyniki dotyczące wydajności mogą się różnić w zależności do konkretnych warunków konfiguracyjnych i operacyjnych.

Informacje dotyczące produktów innych podmiotów niż IBM zostały uzyskane od dostawców tych produktów, z ich publicznych ogłoszeń lub innych dostępnych publicznie źródeł. IBM nie testował tych produktów i nie może potwierdzić dokładności pomiarów wydajności, kompatybilności ani żadnych innych danych związanych z produktami firm innych niż IBM. Pytania dotyczące możliwości produktów firm innych niż IBM należy kierować do dostawców tych produktów.

Wszelkie stwierdzenia dotyczące przyszłych kierunków rozwoju i zamierzeń IBM mogą zostać zmienione lub wycofane bez powiadomienia.

Publikacja ta zawiera przykładowe dane i raporty używane w codziennych operacjach działalności gospodarczej. W celu kompleksowego zilustrowania tej działalności podane przykłady zawierają nazwy osób, firm i ich produktów. Wszystkie te nazwy/nazwiska są fikcyjne i jakiekolwiek podobieństwo do istniejących nazw/nazwisk jest całkowicie przypadkowe.

LICENCJA W ZAKRESIE PRAW AUTORSKICH:

Niniejsza publikacja zawiera przykładowe aplikacje w kodzie źródłowym ilustrujące techniki programowania w różnych systemach operacyjnych. Użytkownik może kopiować, modyfikować i rozpowszechniać te programy przykładowe w dowolnej formie bez uiszczania opłat na rzecz IBM, w celu rozbudowy, użytkowania, handlowym lub w celu rozpowszechniania aplikacji zgodnych z aplikacyjnym interfejsem programowym dla tego systemu operacyjnego, dla którego napisane były programy przykładowe. Programy przykładowe nie zostały gruntownie przetestowane. IBM nie może zatem gwarantować ani sugerować niezawodności, użyteczności i funkcjonalności tych programów. Programy przykładowe są dostarczane w stanie, w jakim się znajdują ("AS IS"), bez jakichkolwiek gwarancji (rękojmię również wyłącza się). IBM nie ponosi odpowiedzialności za jakiekolwiek szkody wynikające z używania programów przykładowych.

Każda kopia programu przykładowego lub jakikolwiek jego fragment, jak też jakiekolwiek prace pochodne muszą zawierać następujące uwagi dotyczące praw autorskich:

Znaki towarowe

IBM, logo IBM i ibm.com są znakami towarowymi lub zastrzeżonymi znakami towarowymi International Business Machines Corp., zarejestrowanymi w wielu systemach prawnych na całym świecie. Pozostałe nazwy produktów i usług mogą być znakami towarowymi IBM lub innych przedsiębiorstw. Aktualna lista znaków towarowych IBM dostępna jest w serwisie WWW, w sekcji "Copyright and trademark

information" (Informacje o prawach autorskich i znakach towarowych), pod adresem www.ibm.com/legal/copytrade.shtml.

Adobe, logo Adobe, PostScript oraz logo PostScript są znakami towarowymi lub zastrzeżonymi znakami towarowymi Adobe Systems Incorporated w Stanach Zjednoczonych i/lub w innych krajach.

Intel, logo Intel, Intel Inside, logo Intel Inside, Intel Centrino, logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium i Pentium są znakami towarowymi lub zastrzeżonymi znakami towarowymi Intel Corporation lub przedsiębiorstw podporządkowanych w Stanach Stanach Zjednoczonych i w innych krajach.

Linux jest zastrzeżonym znakiem towarowym Linusa Torvaldsa w Stanach Zjednoczonych i/lub w innych krajach.

Microsoft, Windows, Windows NT oraz logo Windows są znakami towarowymi Microsoft Corporation w Stanach Zjednoczonych i/lub w innych krajach.

UNIX jest zastrzeżonym znakiem towarowym Open Group w Stanach Zjednoczonych i w innych krajach.

Java oraz wszystkie znaki towarowe i logo dotyczące Java są znakami towarowymi firmy i jej firm zależnych.

Indeks

S

T

U

W

Z