

*IBM SPSS Decision Trees 29*



**Poznámka**

Před použitím těchto informací a produktu, který podporují, si přečtěte informace v tématu “Upozornění” na stránce 21.

**Informace o produktu**

Toto vydání se vztahuje k verzi 29, vydání 0, modifikaci 1 produktu IBM® SPSS Statistics a ke všem následujícím vydáním a modifikacím, dokud nebude v nových vydáních uvedeno jinak.

© Copyright International Business Machines Corporation .

---

# Obsah

<b>Kapitola 1. Rozhodovací stromy.....</b>	<b>1</b>
Vytvoření rozhodnutí Trees.....	1
Výběr kategorií.....	3
Ověřování.....	4
Strom-Rostoucí kritéria.....	5
Volby.....	8
Ukládání informací o modelu.....	11
Výstup.....	11
Editor stromu.....	15
Práce s velkými stromy.....	16
Řízení informací zobrazených ve stromu.....	17
Změna barev stromů a textových písem.....	17
Výběr případu a pravidla přidělení skóre.....	18
<b>Upozornění.....</b>	<b>21</b>
Ochranné známky.....	22
<b>Rejstřík.....</b>	<b>25</b>



---

# Kapitola 1. Rozhodovací stromy

Následující funkce rozhodování stromu jsou obsaženy v části SPSS Statistics Professional Edition nebo volba Rozhodovací nástroje.

## Vytvoření rozhodnutí Trees

---

Procedura stromu rozhodnutí vytváří model klasifikace založený na stromu. Rozdělí případy do skupin nebo předpovídá hodnoty závislé (cílové) proměnné založené na hodnotách nezávislých (prediktorů) proměnných. Tento postup poskytuje ověřovací nástroje pro účely analýzy a konfirmační analýzy klasifikace.

Postup lze použít pro:

**Segmentace.** Identifikujte osoby, které pravděpodobně budou členy určité skupiny.

**Stratifikace.** Přiřadte případy do jedné z několika kategorií, jako jsou například skupiny s vysokou, střední a nízkou rizikovými skupinami.

**Předpověď.** Vytvářejte pravidla a používejte je k předpovídání budoucích událostí, jako je pravděpodobnost, že někdo bude přednastaven na půjčku nebo potenciální hodnotu dalšího prodeje ve vozidle nebo v domácnosti.

**Redukce dat a detekční kontroly proměnných.** Vyberte užitečnou podmnožinu prediktorů z velké sady proměnných pro použití při sestavování formálního parametrického modelu.

**Identifikace interakce.** Identifikujte vztahy, které se týkají pouze specifických podskupin, a zadejte je v rámci formálního parametrického modelu.

**Spojování kategorií a diskrétnost spojitých proměnných.** Překóduje kategorie prediktorů skupin a průběžné proměnné s minimálními ztrátami informací.

**Příklad.** Banka chce úvěrové záruky kategorizovat podle toho, zda představují přiměřené úvěrové riziko, či nikoli. Na základě různých faktorů, včetně známých ratingů minulých zákazníků, můžete sestavit model, který předpovídá, zda budoucí zákazníci budou pravděpodobně na svých úvěrech platební neschopnost.

Analýza na bázi stromu nabízí některé atraktivní funkce:

- Umožňuje identifikovat homogenní skupiny s vysokým nebo nízkým rizikem.
- Díky tomu lze snadno vytvářet pravidla pro vytváření předpovědí o jednotlivých případech.

Aspekty dat

**Data.** Závislé a nezávislé proměnné mohou být:

- *Nominální.* Proměnná může být považována za nominální, když její hodnoty představují kategorie bez vlastního hodnocení (například oddělení společnosti, v níž zaměstnanec pracuje). Příklady nominálních proměnných zahrnují region, poštovní směrovací číslo a náboženskou příslušnost.
- *Pořadové.* Proměnnou lze považovat za ordinální, když její hodnoty představují kategorie s nějakým vlastním hodnocením (například úroveň spokojenosti služeb od vysoce nespokojené s vysoce spokojené). Příklady ordinálních proměnných zahrnují skóre přístupu představující stupeň spokojenosti nebo důvěryhodnost a skóre hodnocení preferencí.
- *Měřítko.* Proměnná může být považována za měřítko (souvislá), když její hodnoty reprezentují uspořádané kategorie se smysluplnou metrikou, takže porovnání vzdálenosti mezi hodnotami je vhodné. Příklady proměnných měřítka zahrnují věk v rocích a příjem v tisících dolarů.

**Frekvence frekvencí** Je-li použita váha váhy, jsou zlomkové váhy zaokrouhleny na nejbližší celé číslo; takže případy s hodnotou váhy menší než 0,5 mají přiřazenu váhu 0, a proto jsou z analýzy vyloučeny.

**Předpoklady.** Tato procedura předpokládá, že byla přiřazena vhodná úroveň měření pro všechny proměnné analýzy, a některé funkce předpokládají, že všechny hodnoty závislé proměnné zahrnuté v analýze mají nadefinované popisky hodnot.

- **Úroveň měření.** Úroveň měření ovlivňuje výpočty stromu, takže všechny proměnné by měly být přiřazeny k odpovídající úrovni měření. Standardně se předpokládá, že číselné proměnné jsou stupnice a proměnné řetězců jsou považovány za nominální, což nemusí přesně odrážet skutečnou úroveň měření. Ikona vedle každé proměnné v seznamu proměnných identifikuje typ proměnné.

Tabulka 1. Ikony úrovně měření

Ikona	Úroveň měření
	Měřitko
	Nominální
	Pořadové

Úroveň měření proměnné můžete dočasně změnit tak, že pravým tlačítkem myši klepnete na proměnnou ze seznamu zdrojových proměnných a vyberete úroveň měření z rozevírací nabídky.

- **Popisky hodnot.** Rozhraní dialogového okna pro tento postup předpokládá, že buď všechny nechybějící hodnoty závislé proměnné kategorické (nominální, ordinální) mají definované popisky hodnot, nebo žádné z nich. Některé funkce nejsou k dispozici, pokud alespoň dvě nechybějící hodnoty kategoriální závislé proměnné mají popisky hodnot. Pokud alespoň dvě nechybějící hodnoty mají nadefinované popisky hodnot, všechny případy s ostatními hodnotami, které nemají popisky hodnot, budou vyloučeny z analýzy.

Jak získat rozhodovací stromy

1. Z nabídky vyberte:

**Analyzovat > Klasifikovat > strom ...**

2. Vyberte závislou proměnnou.
3. Vyberte jednu nebo více nezávislých proměnných.
4. Vyberte rostoucí metodu.

Volitelně můžete:

- Změňte úroveň měření pro jakoukoli proměnnou ve zdrojovém seznamu.
- Vynutíte první proměnnou v seznamu nezávislých proměnných do modelu jako první dělenou proměnnou.
- Vyberte proměnnou ovlivnění, která definuje, jak velký vliv má případ ve stromu-rostoucí proces. Případy s nižšími hodnotami vlivu mají menší vliv; případy s vyššími hodnotami mají větší počet. Hodnoty proměnných vlivu musí být kladné.
- Ověřte strom.
- Přizpůsobte si stromová kritéria.
- Ukládat čísla uzlů terminálu, předpovídané hodnoty a předpovídané pravděpodobnosti jako proměnné.
- Uložte model ve formátu XML (PMML).

Pole s neznámou úrovní měření

Výstraha na úrovni měření se zobrazí, když je úroveň měření pro jednu nebo více proměnných (polí) v datové sadě neznámá. Jelikož úroveň měření ovlivňuje výpočet výsledků pro tuto proceduru, všechny proměnné musí mít definovanou úroveň měření.

**Data skenování.** Přečte data v aktivní datové sadě a přiřadí výchozí úroveň měření k jakýmkoli polím s momentálně neznámou úrovní měření. Je-li datová sada velká, může to nějakou dobu trvat.

**Přiřadit ručně.** Otevře dialogové okno se seznamem všech polí s neznámou úrovní měření. Toto dialogové okno můžete použít k přiřazení úrovně měření k těmto polím. Úroveň měření můžete také přiřadit v pohledu Proměnné v editoru dat.

Vzhledem k tomu, že úroveň měření je pro tuto proceduru důležitá, nemůžete přistupovat k dialogovému oknu pro spuštění této procedury, dokud nebude mít všechna pole definovanou úroveň měření.

Změna úrovně měření

1. Klepněte pravým tlačítkem myši na proměnnou ve zdrojovém seznamu.
2. Vyberte úroveň měření z rozevírací nabídky.

Tím se změní úroveň měření dočasně pro použití v proceduře rozhodovacího stromu.

Metody rostoucích

Dostupné pěstitelské metody jsou:

**CHAID.** Chi-druhá detekce interakce s automatickou interakcí. V každém kroku volba CHAID volí nezávislou proměnnou (prediktor), která má nejsilnější interakci se závislou proměnnou. Kategorie jednotlivých prediktorů se sloučí, pokud se výrazně neliší s ohledem na závislou proměnnou.

**Vyčerpávající CHAID.** Úprava CHAID, která prozkoumá všechny možné rozdělení pro každý prediktor.

**CRT.** Klasifikační a regresní stromy. CRT rozdělí data na segmenty, které jsou co nejvíce homogenní s ohledem na závislou proměnnou. Uzel terminálu, v němž všechny případy mají stejnou hodnotu pro závislou proměnnou, je homogenní, "čistý" uzel.

**QUEST.** Rychlý, nezaujatý, efektivní statistický strom. Metoda, která je rychlá a vyhýbá se jiným metodám, se upřednostňuje prediktory s mnoha kategoriemi. QUEST lze zadat pouze tehdy, je-li závislá proměnná nominální.

S každou metodou jsou výhody a omezení, včetně:

Funkce	CHAID *	CRT	QUEST
Chi-square-based**	X		
Náhradní proměnné nezávislé (prediktory)		X	X
Smazání stromů		X	X
Rozdělení vícecestného uzlu	X		
Dělení binárního uzlu		X	X
Ovlivňující proměnné	X	X	
Předchozí pravděpodobnosti		X	X
Různé náklady na klasifikaci	X	X	X
Rychlý výpočet	X		X

\*Zahrnuje Vyčerpávající CHAID.

\*\* QUEST také používá ukazatel chí kvadrát pro nominální nezávislé proměnné.

## Výběr kategorií

Pro kategoriální (nominální, ordinální) závislé proměnné můžete provádět následující akce:

- Řídí, které kategorie jsou zahrnuty v analýze.

- Identifikujte cílové kategorie, které vás zajímají.

Zahrnutí/vyloučení kategorií

Analýzu můžete omezit na specifické kategorie závislé proměnné.

- Případy s hodnotami závislé proměnné v seznamu Vyloučit nejsou zahrnuty do analýzy.
- U nominálních závislých proměnných můžete do analýzy zahrnout také uživatele-chybějící kategorie. (Ve výchozím nastavení se v seznamu Vyloučit zobrazují kategorie chybějících kategorií.)

Cílové kategorie

Vybrané (kontrolované) kategorie se považují za kategorie primárního zájmu v analýze. Například pokud se primárně zajímáte o identifikaci osob s největší pravděpodobností k selhání v půjčce, můžete jako cílovou kategorii vybrat "nesprávnou" kategorii úvěru-hodnocení.

- Neexistuje žádná výchozí cílová kategorie. Není-li vybrána žádná kategorie, nejsou k dispozici některé volby klasifikačních pravidel a výstupy související se zisky.
- Je-li vybráno více kategorií, jsou pro každou cílovou kategorii vytvořeny samostatné tabulky zisků a grafy.
- Označení jedné nebo více kategorií jako cílových kategorií nemá žádný vliv na model stromu, odhad rizika nebo nesprávné výsledky klasifikace.

Popisky kategorií a hodnot

Toto dialogové okno vyžaduje definované popisky hodnot pro závislou proměnnou. Není k dispozici, pokud alespoň dvě hodnoty kategoriální závislé proměnné nemají definované popisky hodnot.

Chcete-li zahrnout/vyloučit kategorie a vybrat cílové kategorie

1. V hlavním dialogovém okně Strom rozhodnutí vyberte kategoriální (nominální, ordinální) závislou proměnnou se dvěma nebo více nadefinovanými popisky hodnot.
2. Klepněte na volbu **Kategorie**.

## Ověřování

Validace vám umožňuje posoudit, jak dobře vaše stromová struktura se zevšeobecňuje na větší populaci. K dispozici jsou dvě metody validace: křížové ověření platnosti a rozdělení-validace vzorků.

Prokuku

Crossvalidation rozděluje vzorek na několik dílčích vzorků nebo **záhyby**. Vygenerují se modelové modely, s vyloučením dat z každého dílčího vzorku na otočku. První strom je založen na všech případech kromě těch v první ukázce, druhý strom je založen na všech případech s výjimkou těch, které jsou uvedeny v druhém vzorku a tak dále. U každého stromu se riziko nesprávné klasifikace odhaduje použitím stromu na dílčí vzorek, který je vyloučen při generování.

- Můžete uvést maximálně 25 vzorkovací záhyby. Čím vyšší je hodnota, tím méně je počet případů vyloučených pro každý model stromu.
- Crossvalidation produkuje jediný, konečný stromový model. Odhad rizika pro konečný strom je vypočítán jako průměr rizik pro všechny stromy.

Rozdělit-Validace vzorku

S validací děleného vzorku se model vygeneruje pomocí cvičovacího vzorku a otestuje se na hold-out vzorku.

- Můžete uvést velikost ukázky školení, vyjádřenou jako procentní část celkové velikosti vzorku nebo proměnné, která rozdělí vzorek na vzorky školení a testování.
- Pokud použijete proměnnou k definování školení a testování ukázek, jsou případy s hodnotou 1 pro danou proměnnou přiřazeny do ukázky školení a všechny ostatní případy jsou přiřazeny k ukázce testování. Proměnná nemůže být závislá proměnná, váhový proměnná, ovlivnění proměnné nebo vynucená nezávislá proměnná.



- Můžete zobrazit výsledky pro vzorky školení i testování nebo pouze ukázkou testování.
- U malých datových souborů (datové soubory s malým počtem případů) by mělo být prováděno ověřování děleného vzorku. Malé velikosti výběrového souboru mohou přinést špatné modely, protože v některých kategoriích nemusí být dostatečný počet případů, které odpovídajícím způsobem rostou ve stromu.

Ověření stromu rozhodnutí

1. V hlavním dialogovém okně Třes rozhodnutí klepněte na tlačítko **Ověření platnosti**.
2. Vyberte volbu **Crossvalidation** nebo **Split-sample validation**.

*Poznámka:* Obě metody ověření náhodně přiřazují případy k vzorkovým skupinám. Chcete-li být schopni reprodukovat přesně stejné výsledky v následné analýze, měli byste před prvním spuštěním analýzy nastavit náhodný počet náhodných čísel (Transform menu, Náhodný počet generátorů) a pak resetujete počáteční hodnotu na tuto hodnotu pro následnou analýzu.

## Strom-Rostoucí kritéria

Dostupná rostoucí kritéria mohou záviset na rostoucí metodě, úrovni měření závislé proměnné nebo na kombinaci obou těchto kritérií.

### Limity růstu

Ouško Omezení růstu vám umožňuje omezit počet úrovní ve stromu a řídit minimální počet případů pro nadřizené a podřizené uzly.

**Maximální hloubka stromu.** Řídí maximální počet úrovní růstu pod kořenovým uzlem. Nastavení **Automatický** omezuje strom na tři úrovně pod kořenovým uzlem pro metody CHAID a Vyčerpávající metody CHAID a pět úrovní pro metody CRT a QUEST.

**Minimální počet případů.** Řídí minimální počet případů pro uzly. Uzly, které nesplňují tato kritéria, nebudou rozděleny.

- Zvětšení minimálních hodnot má tendenci produkovat stromy s menším počtem uzlů.
- Zmenšením minimálních hodnot se vytvoří stromy s více uzly.

U datových souborů s malým počtem případů mohou výchozí hodnoty 100 případů pro nadřizené uzly a 50 případů pro podřizené uzly někdy vést ke stromům, které nemají pod kořenovým uzlem žádné uzly. V takovém případě může snížení minimální hodnoty přinést více užitečných výsledků.

Jak specifikovat mezní hodnoty růstu

1. V hlavním dialogovém okně Strom rozhodnutí klepněte na volbu **Kritéria**.
2. Klepněte na kartu **Limity růstu**.

### Kritéria CHAID

Pro metody CHAID a Vyčerpávající metody CHAID můžete řídit:

**Úroveň významnosti.** Můžete určit hodnotu významnosti pro rozdělení uzlů a slučování kategorií. Pro obě kritéria je výchozí úroveň významnosti 0,05.

- Pro rozdělování uzlů musí být hodnota větší než 0 a menší než 1. Nižší hodnoty mají tendenci vytvářet stromy s menším počtem uzlů.
- Pro slučování kategorií musí být hodnota větší než 0 a menší než nebo rovna 1. Chcete-li zabránit slučování kategorií, zadejte hodnotu 1. Pro měřítko nezávislé proměnné to znamená, že počet kategorií pro proměnnou v konečném stromu je určený počet intervalů (výchozí nastavení je 10). Další informace naleznete v tématu [“Intervaly stupnice pro analýzu CHAID”](#) na stránce 6.

**Chi-Square Statistic.** Pro ordinální závislé proměnné se hodnota chí-kvadrát pro určení rozdělení uzlu a sloučení kategorie vypočítá pomocí metody poměru pravděpodobnosti. Pro nominální závislé proměnné můžete vybrat metodu:

- **Pearson.** Tato metoda poskytuje rychlejší výpočty, ale měla by být používána s opatrností u malých vzorků. Jedná se o výchozí metodu.
- **Poměr pravděpodobnosti.** Tato metoda je mnohem robustnější, než Pearson, ale trvá déle, než vypočítat. U malých vzorků se jedná o upřednostňovanou metodu.

**Odhad modelu.** Pro nominální a pořadové závislé proměnné můžete zadat:

- **Maximální počet iterací.** Výchozí hodnota je 100. Pokud strom přestane růst, protože byl dosažen maximální počet iterací, můžete chtít zvýšit maximum nebo změnit jedno nebo více dalších kritérií, která ovládají růst stromu.
- **Minimální změna v očekávaných frekvencích buněk.** Hodnota musí být větší než 0 a menší než 1. Výchozí hodnota je 0,05. Nižší hodnoty mají tendenci vytvářet stromy s menším počtem uzlů.

**Upravte hodnoty významnosti pomocí metody Bonferroni.** Pro více porovnání jsou hodnoty významnosti pro slučování a rozdělování kritérií upraveny pomocí metody Bonferroni. Jedná se o výchozí nastavení.

**Povolit opakované odesílání sloučených kategorií v rámci uzlu.** Pokud explicitně nezabráníte slučování kategorií, procedura se pokusí sloučit nezávislé skupiny proměnných (prediktory) dohromady a vytvořit tak nejjednodušší strom, který tento model popisuje. Tato volba umožňuje postup opětovného odeslání sloučených kategorií, pokud tento způsob poskytuje lepší řešení.

Uvést kritéria CHAID

1. V hlavním dialogovém okně Strom rozhodnutí vyberte jako rostoucí metodu volbu **CHAID** nebo **Vyčerpávající CHAID**.
2. Klepněte na volbu **Kritéria**.
3. Klepněte na kartu **CHAID**.

### **Intervaly stupnice pro analýzu CHAID**

V analýze CHAID jsou proměnné nezávislé (prediktory) vždy sbalené do samostatných skupin (například 0-10, 11-20, 21-30 atd.) před analýzou. Můžete řídit počáteční/maximální počet skupin (ačkoli může procedura slučovat sousedící skupiny po počátečním rozdělení):

- **Pevné číslo.** Všechny nezávislé proměnné stupnice jsou na počátku sbalené do stejného počtu skupin. Výchozí hodnota je 10.
- **Vlastní.** Každá škálovaná proměnná je na počátku rozdělena do počtu skupin určených pro danou proměnnou.

Chcete-li určit intervaly pro měřítko nezávislých proměnných,

1. V hlavním dialogovém okně Strom rozhodnutí vyberte jednu nebo více nezávislých proměnných měřítko.
2. Pro rostoucí metodu vyberte **CHAID** nebo **Vyčerpávající CHAID**.
3. Klepněte na volbu **Kritéria**.
4. Klepněte na kartu **Intervaly**.

V CRT a QUEST analýzy jsou všechny rozdělení binární a stupnice a ordinální nezávislé proměnné jsou zpracovány stejným způsobem; takže nemůžete uvést počet intervalů pro měřítko nezávislých proměnných.

### **Kritéria CRT**

Metoda CRT rostoucí metodou se snaží maximalizovat v rámci-uzel homogenity. Oblast, v níž uzel nereprezentuje homogenní podmnožinu případů, je údaj o **nečistotu**. Např. uzel terminálu, ve kterém všechny případy mají stejnou hodnotu pro závislou proměnnou je homogenní uzel, který nevyžaduje žádné další rozdělování, protože je "čistý".

Můžete vybrat metodu použitou k měření nečistoty a minimální snížení nečistoty potřebné k rozdělení uzlů.

**Ukazatel nečistoty.** U proměnných závislých na měřítku se použije ukazatel nejnižší odchylky (LSD), který je nečistotou. Vypočítá se jako odchylka v rámci uzlu, upravená pro všechny frekvenční váhy nebo hodnoty vlivu.

Pro kategoriální (nominální, ordinální) proměnné závislé proměnné můžete vybrat nečistotu:

- **Gini.** Rozděluje se zjistí, že se maximalizuje homogenita podřízených uzlů s ohledem na hodnotu závislé proměnné. Gini je založena na kvadrání pravděpodobnosti členství pro každou kategorii závislé proměnné. Dosáhne svého minima (nula), když všechny případy v uzlu spadají do jedné kategorie. Toto je výchozí ukazatel.
- **Twoing.** Kategorie závislé proměnné jsou seskupeny do dvou podtříd. Splits jsou nalezeny, že nejlépe oddělit dvě skupiny.
- **Uspořádáno dvě informace.** Podobně jako dvě, kromě toho, že lze seskupit pouze sousední kategorie. Toto měřítko je k dispozici pouze pro ordinální závislé proměnné.

**Minimální změna ve zlepšení.** Jedná se o minimální snížení nečistoty potřebné k rozdělení uzlu. Výchozí hodnota je 0, 0001. Vyšší hodnoty mají tendenci vytvářet stromy s menším počtem uzlů.

Chcete-li zadat kritéria CRT

1. Pro rostoucí metodu vyberte **CRT**.
2. Klepněte na volbu **Kritéria**.
3. Klepněte na kartu **CRT**.

## Kritéria QUEST

Pro metodu QUEST můžete určit úroveň významnosti pro rozdělování uzlů. Nezávislá proměnná nemůže být použita k rozdělení uzlů, pokud úroveň významnosti není menší než nebo rovna zadané hodnotě. Hodnota musí být větší než 0 a menší než 1. Výchozí hodnota je 0,05. Menší hodnoty budou mít tendenci vyloučit více nezávislých proměnných z konečného modelu.

Chcete-li určit kritéria QUEST

1. V hlavním dialogovém okně Strom rozhodnutí vyberte nominální závislou proměnnou.
2. Pro rostoucí metodu vyberte **QUEST**.
3. Klepněte na volbu **Kritéria**.
4. Klepněte na kartu **QUEST**.

## Stromy stromů

S metodami CRT a QUEST se můžete vyhnout překrytí modelu pomocí **pročišťování** stromu: strom se rozrůstá, dokud nebudou splněna kritéria, a pak se automaticky ořízne na nejmenší podstrom na základě uvedeného maximálního rozdílu v riziku. Hodnota rizika je vyjádřena ve standardních chybách. Výchozí nastavení je 1. Hodnota musí být nezáporná. Chcete-li získat podstrom s minimálním rizikem, uveďte 0.

Smazat strom

1. V hlavním dialogovém okně Strom rozhodování pro rostoucí metodu vyberte **CRT** nebo **QUEST**.
2. Klepněte na volbu **Kritéria**.
3. Klepněte na kartu **Smazání**.

Smazání a skrytí uzlů

Když vytvoříte smazaný strom, všechny uzly smazané ze stromu nebudou v konečném stromu k dispozici. Můžete interaktivně skrýt a zobrazit vybrané podřízené uzly v posledním stromě, ale nemůžete zobrazit uzly, které byly vyčištěny v procesu vytváření stromu. Další informace naleznete v tématu [“Editor stromu”](#) na stránce 15 .

## Náhradníci

CRT a QUEST může použít **surrogates** pro nezávislé proměnné (prediktory). Pro případy, ve kterých hodnota proměnné chybí, jsou pro klasifikaci použity jiné nezávislé proměnné s velkými přidruženými s původní proměnnou. Tyto alternativní prediktory se nazývají surrogates. Můžete určit maximální počet náhradníků, které se použijí v modelu.

- Při výchozím nastavení je maximální počet náhradních uživatelů o jeden menší než počet nezávislých proměnných. Jinými slovy, pro každou nezávislou proměnnou mohou být jako náhradníky použity všechny ostatní nezávislé proměnné.
- Pokud nechcete, aby model použil surrogates, uveďte 0 pro počet náhradníci.

Chcete-li určit náhradníky

1. V hlavním dialogovém okně Strom rozhodování pro rostoucí metodu vyberte **CRT** nebo **QUEST**.
2. Klepněte na volbu **Kritéria**.
3. Klepněte na kartu **Surrogates**.

## Volby

Dostupné volby mohou záviset na rostoucí metodě, úrovni měření závislé proměnné a/nebo existenci popisků hodnot definovaných hodnot pro hodnoty závislé proměnné.

### Různé náklady na klasifikaci

Pro závislé proměnné pro kategoriální (nominální, ordinální) umožňují nesprávné klasifikace nákladů zahrnout informace o relativní pokutě přidružené k nesprávné klasifikaci. Příklad:

- Náklady na odepření úvěru úvěruschopnému zákazníkovi se pravděpodobně liší od nákladů na rozšíření úvěru na zákazníka, který pak bude splácet úvěr.
- Náklady na nesprávné zařazení jednotlivce s vysokým rizikem onemocnění srdce jako nízké riziko jsou pravděpodobně mnohem vyšší než náklady na zpronevěra vysoce rizikové jedince jako vysoce rizikové.
- Náklady na odeslání hromadné pošty někomu, kdo pravděpodobně neodpoví, bude pravděpodobně poměrně nízké, zatímco náklady na zaslání pošty někomu, kdo pravděpodobně odpoví, jsou relativně vyšší (z hlediska ušlého příjmu).

Nákladovost neklasifikace a popisky hodnot

Toto dialogové okno není k dispozici, pokud alespoň dvě hodnoty z kategoriální závislé proměnné nemají definované popisky hodnot.

Uvedení nákladů na klasifikaci

1. V hlavním dialogovém okně Strom rozhodnutí vyberte kategoriální (nominální, ordinální) závislou proměnnou se dvěma nebo více nadefinovanými popisky hodnot.
2. Klepněte na volbu **Volby**.
3. Klepněte na kartu **Náklady na klasifikaci**.
4. Klepněte na **Vlastní**.
5. Zadejte jednu nebo více neklasifikačních nákladů v mřížce. Hodnoty musí být nezáporné. (Správné klasifikace, představované na úhlopříčce, jsou vždy 0.)

**Vyplnit matici.** V mnoha případech můžete chtít, aby byly náklady symetrické-to znamená, že náklady na chybné klasifikace A jako B jsou stejné jako náklady na chybné klasifikaci B jako A. Pomocí následujících ovládacích prvků lze snáze určit matici symetrických nákladů:

- **Duplikovat dolní trojúhelník.** Kopíruje hodnoty do spodního trojúhelníku matice (pod úhlopříčkou) do odpovídajících horních trojúhelníkových buněk.
- **Duplikovat horní trojúhelník.** Zkopíruje hodnoty do horního trojúhelníku matice (nad úhlopříčkou) do odpovídajících spodních trojúhelníkových buněk.

- **Použit průměrné hodnoty buněk.** Pro každou buňku v každé polovině matice jsou obě hodnoty (horní a dolní trojúhelníkové) zprůměrovány a průměrná hodnota nahradí obě hodnoty. Je-li například cena chybné klasifikace A jako B 1 a cena chybné klasifikace B jako A je 3, pak tento ovládací prvek nahradí obě tyto hodnoty průměrem  $(1 + 3)/2 = 2$ .

## Zisk

U kategoriálních závislých proměnných můžete přiřadit hodnoty výnosů a výdajů k úrovním závislé proměnné.

- Zisk se vypočítá jako výnos minus výdaj.
- Hodnoty zisku mají vliv na průměrné hodnoty zisku a návratnosti investic (návratnost investic) v tabulkách zisků. Nemají vliv na základní strukturu modelu stromu.
- Hodnoty výnosů a výdajů musí být číselné a musí být uvedeny pro všechny kategorie závislé proměnné zobrazené v mřížce.

Popisky vlastností a hodnot

Toto dialogové okno vyžaduje definované popisky hodnot pro závislou proměnnou. Není k dispozici, pokud alespoň dvě hodnoty kategoriální závislé proměnné nemají definované popisky hodnot.

Chcete-li zadat zisky

1. V hlavním dialogovém okně Strom rozhodnutí vyberte kategoriální (nominální, ordinální) závislou proměnnou se dvěma nebo více nadefinovanými popisky hodnot.
2. Klepněte na volbu **Volby**.
3. Klepněte na kartu **Zisky**.
4. Klepněte na **Vlastní**.
5. Zadejte hodnoty výnosů a výdajů pro všechny kategorie závislé proměnné uvedené v mřížce.

## Předchozí pravděpodobnosti

Pro stromy CRT a QUEST s kategoriálními závislými proměnnými můžete určit předchozí pravděpodobnosti členství ve skupině. **Předchozí pravděpodobnosti** jsou odhady celkové relativní četnosti pro každou kategorii závislé proměnné před tím, než se dozví cokoli o hodnotách nezávislých (prediktorů) proměnných. Použití předchozích pravděpodobností pomáhá opravit jakýkoli růst stromu způsobený daty ve vzorku, která není reprezentativní pro celou populaci.

**Obstarejte si ze vzorku školení (empirické primáry).** Toto nastavení použijte, pokud je distribuce závislých hodnot proměnných v datovém souboru reprezentativní pro distribuci populace. Používáte-li rozdělení-ukázkové ověření, použijte se distribuce případů ve vzorku školení.

*Poznámka:* Vzhledem k tomu, že se případy náhodně přiřazují do vzorku školení v rámci validace děleného vzorku, nebudete znát skutečnou distribuci případů předem ve vzorku školení. Další informace naleznete v tématu [“Ověřování”](#) na stránce 4.

**Rovno mezi kategoriemi.** Toto nastavení použijte v případě, že kategorie závislé proměnné jsou rovnoměrně zastoupeni v populaci. Jsou-li například v každé kategorii čtyři kategorie, přibližně 25% případů je v každé kategorii.

**Vlastní.** Zadejte nezápornou hodnotu pro každou kategorii závislé proměnné uvedené v mřížce. Hodnoty mohou být proporcionální, procentní části, frekvence nebo jakékoli jiné hodnoty, které představují rozdělení hodnot v různých kategoriích.

**Upravte záznamy za použití neklasifikačních nákladů.** Definujete-li vlastní náklady na nesprávné klasifikace, můžete upravit předchozí pravděpodobnosti na základě těchto nákladů. Další informace naleznete v tématu [“Různé náklady na klasifikaci”](#) na stránce 8.

Popisky vlastností a hodnot

Toto dialogové okno vyžaduje definované popisky hodnot pro závislou proměnnou. Není k dispozici, pokud alespoň dvě hodnoty kategoriální závislé proměnné nemají definované popisky hodnot.

Chcete-li zadat předchozí pravděpodobnosti

1. V hlavním dialogovém okně Strom rozhodnutí vyberte kategoriální (nominální, ordinální) závislou proměnnou se dvěma nebo více nadefinovanými popisky hodnot.
2. Pro rostoucí metodu vyberte **CRT** nebo **QUEST**.
3. Klepněte na volbu **Volby**.
4. Klepněte na kartu **Předchozí pravděpodobnosti**.

## Skóre

Pro CHAID a Vyčerpávající CHAID s ordinální závislou proměnnou můžete přiřadit vlastní skóre ke každé kategorii závislé proměnné. Skóre definuje pořadí a vzdálenost mezi kategoriemi závislé proměnné. Můžete použít skóre ke zvýšení nebo snížení relativní vzdálenosti mezi pořadovými hodnotami nebo ke změně pořadí hodnot.

- **Pro každou kategorii použijte pořadové číslo.** Nejnižší kategorie závislé proměnné má přiřazeno skóre 1, další nejvyšší kategorie má přiřazeno skóre 2 atd. Jedná se o výchozí nastavení.
- **Vlastní.** Zadejte číselnou hodnotu skóre pro každou kategorii závislé proměnné uvedené v mřížce.

Příklad

Popisek hodnoty	Původní hodnota	Skóre
Nekvalifikované	1	1
Příručka Shilled	2	4
Úředník	3	4.5
Profesionální	4	7
správa	5	6

- Skóre zvýší relativní vzdálenost mezi *Nekvalifikovaní* a *ruční ruční* a sníží relativní vzdálenost mezi *ručním ručním systémem* a *Cerical*.
- Skóre inverzní k pořadí *Správa* a *Profesionální*.

Popisky skóre a hodnot

Toto dialogové okno vyžaduje definované popisky hodnot pro závislou proměnnou. Není k dispozici, pokud alespoň dvě hodnoty kategoriální závislé proměnné nemají definované popisky hodnot.

Chcete-li určit skóre

1. V hlavním dialogovém okně Strom rozhodnutí vyberte ordinální závislou proměnnou se dvěma nebo více nadefinovanými popisky hodnot.
2. Pro rostoucí metodu vyberte **CHAID** nebo **Vyčerpávající CHAID**.
3. Klepněte na volbu **Volby**.
4. Klepněte na kartu **Scores**.

## Chybějící hodnoty

Ouško Chybějící hodnoty řídí obsluhu nominálních hodnot proměnných, které chybí, jsou nezávislé (prediktor).

- Zpracování ordinálních a škálovaných uživatelů-chybějící hodnoty nezávislé proměnné se liší mezi rostoucími metodami.
- Práce s nominálními závislými proměnnými je uvedena v dialogovém okně Kategorie. Další informace naleznete v tématu [“Výběr kategorií”](#) na stránce 3.

- Pro ordinální a měřítko závislé proměnné jsou vždy vyloučeny případy se systémem chybějícími nebo uživateli-chybějící závislé hodnoty proměnných jsou vždy vyloučeny.

**Považovat za chybějící hodnoty.** Hodnoty uživatelských chybějících hodnot jsou považovány za systémové chyběly hodnoty. Způsob zacházení se systémem/chybějící hodnoty se liší podle rostoucí metody.

**Považovat za platné hodnoty.** Uživatelské hodnoty nominálních nezávislých proměnných jsou považovány za běžné hodnoty ve stromovém pěstování a klasifikaci.

Metoda-závislá pravidla

Pokud některé, ale ne všechny, nezávislé proměnné hodnoty jsou systémem nebo uživatelem-chybí:

- Pro CHAID a Vyčerpávající CHAID jsou systémové hodnoty a hodnoty nezávislé na uživateli zahrnuty v analýze jako jedna kombinovaná kategorie. Pro měřítko a ordinální nezávislé proměnné algoritmy nejprve vygenerují kategorie pomocí platných hodnot a poté se rozhodnou, zda sloučit chybějící kategorii s největší podobnou (platnou) kategorií, nebo zachovat ji jako samostatnou kategorii.
- Pro CRT a QUEST jsou případy s chybějícími nezávislými variabilními hodnotami vyloučeny ze stromového procesu, ale jsou klasifikovány pomocí náhradníků, jsou-li náhrady zahrnuty do metody. Pokud se s nominálním uživatelem chybějící hodnoty zachází jako s chybějícími hodnotami, jsou s nimi také manipulovány tímto způsobem. Další informace naleznete v tématu [“Náhradníci” na stránce 8](#).

Chcete-li zadat nominální, nezávislý uživatel-chybějící zpracování

1. V hlavním dialogovém okně Strom rozhodnutí vyberte alespoň jednu nominální nezávislou proměnnou.
2. Klepněte na volbu **Volby**.
3. Klepněte na kartu **Chybějící hodnoty**.

## Ukládání informací o modelu

Informace z modelu můžete uložit jako proměnné do pracovního datového souboru a můžete také uložit celý model ve formátu XML (PMML) do externího souboru.

Uložené proměnné

**Číslo uzlu terminálu.** Uzel terminálu, ke kterému je každý případ přiřazen. Hodnota je číslo uzlu stromu.

**Předpovězená hodnota.** Třída (skupina) nebo hodnota pro závislou proměnnou předpovězenou modelem.

**Předpokládané pravděpodobnosti.** Pravděpodobnost přidružená k předpovědi modelu. Jedna proměnná je uložena pro každou kategorii závislé proměnné. Nejsou k dispozici pro měřítko závislých proměnných.

**Ukázka přiřazení (školení/testování).** U rozdělení-ukázkové ověření tato proměnná označuje, zda byl případ použit ve zkušební nebo testovací ukázce. Hodnota je 1 pro vzorek školení a 0 pro vzorek testování. Není k dispozici, pokud jste nevybrali volbu rozdělení-ukázka ověření. Další informace naleznete v tématu [“Ověřování” na stránce 4](#).

Exportovat stromový model jako XML

Celý stromový model můžete uložit ve formátu XML (PMML). Tento modelový soubor můžete použít k použití informací modelu na jiné datové soubory pro účely hodnocení.

**Ukázka školení.** Zapiše model do zadaného souboru. U dělených vzorků validovaných stromů se jedná o vzor pro odběr vzorků.

**Testovací ukázka.** Zapiše model pro testovací vzorek do zadaného souboru. Není k dispozici, pokud jste nevybrali volbu rozdělení-ukázka ověření.

## Výstup

Dostupné možnosti výstupu závisejí na rostoucí metodě, na úrovni měření závislé proměnné a na dalších nastaveních.

## Zobrazení stromu

Můžete určit počáteční vzhled stromu nebo zcela potlačit zobrazení stromu.

**Strom.** Ve výchozím nastavení je diagram stromu zahrnut do výstupu zobrazeného v prohlížeči. Chcete-li z výstupu vyloučit diagram stromu, zrušte výběr této volby.

**Obrazovka.** Tyto volby řídí počáteční vzhled stromového diagramu v prohlížeči. Všechny tyto atributy lze také upravit úpravou vygenerovaného stromu.

- **Orientace.** Strom může být zobrazen shora dolů s kořenovým uzlem na horním okraji, zleva doprava nebo zprava doleva.
- **Obsah uzlu.** Uzly mohou zobrazovat tabulky, grafy nebo obojí. Pro kategoriální závislé proměnné tabulky zobrazují počty četnosti a procentní části a grafy jsou pruhové grafy. Pro závislé proměnné měřítka tabulky zobrazují standardní odchylky, počet případů a predikované hodnoty a grafy jsou histogramy.
- **Měřítko.** Ve výchozím nastavení jsou velké stromy automaticky zmenšeny v pokusu o přizpůsobení stromu na stránce. Můžete zadat vlastní procento až 200%.
- **Nezávislá statistika proměnných.** Pro CHAID a Vyčerpávající CHAID statistiky zahrnují hodnotu  $F$  (pro proměnné závislé na měřítku) nebo hodnotu chí kvadrát (pro kategoriální závislé proměnné) a také hodnotu významnosti a stupně volnosti. Pro CRT je zobrazena hodnota zlepšení. Pro QUEST,  $F$ , hodnotu významnosti a stupně volnosti jsou zobrazeny pro měřítko a ordinální nezávislé proměnné; pro nominální nezávislé proměnné, je zobrazeno chí kvadrát, hodnota významnosti a stupně volnosti.
- **Definice uzlů.** Definice uzlů zobrazují hodnotu (hodnoty) nezávislé proměnné použité při každém rozdělení uzlu.

**Strom v tabulkovém formátu.** Souhrnné informace pro každý uzel ve stromu, včetně čísla nadřazeného uzlu, statistiky nezávislé proměnné, hodnoty nezávislé proměnné pro uzel, střední a směrodatné odchylky pro proměnné závislé na měřítku, nebo počty a procentní části pro kategoriální závislé proměnné.

Chcete-li řídit zobrazení počátečního stromu

1. V hlavním dialogovém okně Rozhodovací strom klepněte na volbu **Výstup**.
2. Klepněte na kartu **Strom**.

## Statistika

Dostupné statistické tabulky závisejí na úrovni měření závislé proměnné, rostoucí metodě a dalších nastaveních.

Model

**Souhrn.** Souhrn obsahuje použitou metodu, proměnné zahrnuté v modelu a uvedené proměnné, ale nejsou zahrnuty do modelu.

**Riziko.** Odhad rizika a jeho standardní chyba. Ukazatel prediktivní přesnosti stromu.

- U kategoriálních závislých proměnných je odhad rizika poměrným podílem případů nesprávně klasifikovaných po úpravě za předchozí pravděpodobnosti a chybné klasifikace.
- U proměnných závislých na měřítku je odhad rizika proměnný v rámci uzlu.

**Tabulka klasifikace.** Pro závislé proměnné kategorická (nominální, ordinální) tato tabulka zobrazuje počet případů klasifikovaných správně a nesprávně klasifikovaných pro každou kategorii závislé proměnné. Nejsou k dispozici pro měřítko závislých proměnných.

**Náklady, předchozí pravděpodobnost, skóre a hodnoty zisku.** U kategoriálních závislých proměnných tato tabulka zobrazuje náklady, předchozí pravděpodobnost, skóre a hodnoty zisku použité v analýze. Nejsou k dispozici pro měřítko závislých proměnných.

Nezávislé proměnné

**Důležitost k modelu.** U rostoucí metody CRT řadí každou nezávislou (predikční) proměnnou podle svého významu pro model. Není k dispozici pro metody QUEST nebo CHAID.



**Náhradníci podle rozdělení.** Pro metody CRT a QUEST rostoucí metody, pokud model zahrnuje surrogates, lists surrogates pro každé štěpení ve stromě. Není k dispozici pro metody CHAID. Další informace naleznete v tématu [“Náhradníci”](#) na stránce 8 .

Výkon uzlu

**Souhrn.** U proměnných závislých na měřítku obsahuje tabulka číslo uzlu, počet případů a střední hodnotu závislé proměnné. U kategoriálních závislých proměnných s definovanými zisky tabulka obsahuje číslo uzlu, počet případů, průměrný zisk a hodnoty návratnosti investic (návratnost investic). Není k dispozici pro kategoriální závislé proměnné bez definovaných zisků. Další informace naleznete v tématu [“Zisk”](#) na stránce 9 .

**Podle cílové kategorie.** U kategorizovaných závislých proměnných s definovanými cílovými kategoriemi obsahuje tabulka procentní zisk, procentní část odezvy a procentní část indexu (výtah) podle uzlu nebo skupiny percentilů. Pro každou cílovou kategorii se vytvoří samostatná tabulka. Není k dispozici pro měřítko závislých proměnných nebo kategoriálních závislých proměnných bez definovaných cílových kategorií. Další informace naleznete v tématu [“Výběr kategorií”](#) na stránce 3 .

**Řádky.** Tabulky výkonu uzlu mohou zobrazit výsledky podle uzlů terminálu, percentil nebo obojího. Pokud vyberete obě tabulky, vytvoří se pro každou cílovou kategorii dvě tabulky. Tabulky percentilů zobrazují kumulativní hodnoty pro každý percentil, na základě pořadí řazení.

**Přírůstek percentilů.** Pro percentil tabulek můžete vybrat percentil přírůstek: 1, 2, 5, 10, 20 nebo 25.

**Zobrazte kumulativní statistiky.** Pro tabulky uzlů terminálu zobrazí další sloupce v každé tabulce s kumulativními výsledky.

Výběr výstupu statistiky

1. V hlavním dialogovém okně Rozhodovací strom klepněte na volbu **Výstup**.
2. Klepněte na kartu **Statistika** .

## Grafy

Dostupné grafy závisejí na úrovni měření závislé proměnné, rostoucí metodě a dalších nastaveních.

**Nezávislá proměnná důležitá pro model.** Vodorovný pruhový graf důležitosti modelu podle nezávislé proměnné (prediktor). K dispozici pouze s metodou CRT rostoucí metodou.

Výkon uzlu

**Získat.** Zisk je procentní podíl celkového počtu případů v cílové kategorii v každém uzlu, vypočítaný jako:  $(\text{cíl uzlu } n / \text{celkový cíl } n) \times 100$ . Graf zisků je čárový graf kumulativních zisků v percentilech, vypočítaný jako:  $(\text{kumulativní cíl percentilu } n / \text{celkový cíl } n) \times 100$ . Pro každou cílovou kategorii se vytvoří samostatný čárový graf. K dispozici pouze pro kategoriální závislé proměnné s definovanými cílovými kategoriemi. Další informace naleznete v tématu [“Výběr kategorií”](#) na stránce 3 .

Graf přírůstků zakresluje stejné hodnoty, které byste měli vidět ve sloupci *Zisk procent* v tabulce výnosů pro percentily, které také nahlásí kumulativní hodnoty.

**Rejstřík.** Index je poměr procentní části odezvy uzlu pro cílovou kategorii v porovnání s celkovým procentem odezvy cílové kategorie pro celý vzorek. Sloupcový graf je spojnicový graf kumulativních hodnot percentilu indexu. K dispozici pouze pro kategoriální závislé proměnné. Kumulativní index percentilu se vypočítá jako:  $(\text{kumulativní procentní část odezvy v procentech} / \text{celková odpověď v procentech}) \times 100$ . Pro každou cílovou kategorii je vytvořen samostatný diagram a cílové kategorie musí být definovány.

Sloupcový graf zakresluje stejné hodnoty, které byste viděli ve sloupci *Index* v přírůstcích po tabulce percentilů.

**Odezva.** Procento případů v uzlu v uvedené cílové kategorii. Graf odezvy je spojnicový graf kumulativní odezvy percentilu, vypočítaný jako:  $(\text{kumulativní cíl percentilu } n / \text{kumulativní celkový percentil } n) \times 100$ . K dispozici pouze pro kategoriální závislé proměnné s definovanými cílovými kategoriemi.

Graf odezvy zakresluje stejné hodnoty, které byste měli vidět ve sloupci *Odezva* v přírůstcích v tabulce percentilů.

**Střední hodnota.** Čárový graf kumulativních průměrných hodnot percentilu pro závislou proměnnou. K dispozici pouze pro měřítko závislých proměnných.

**Průměrný zisk.** Čárový graf kumulativního průměrného zisku. K dispozici pouze pro kategoriální závislé proměnné s definovanými zisky. Další informace naleznete v tématu [“Zisk”](#) na stránce 9 .

Graf průměru zisku zakresluje stejné hodnoty, které byste měli vidět ve sloupci *Zisk* , v tabulce zisku souhrnu pro percentily.

**Návratnost investic (ROI).** Čárový graf kumulativní ROI (návrátnost investic). Návratnost investic se vypočítá jako poměr zisků k výdajům. K dispozici pouze pro kategoriální závislé proměnné s definovanými zisky.

Graf ROI zakresluje stejné hodnoty, které byste měli vidět ve sloupci *ROI* v tabulce percentilů přehledu pro percentily.

**Přírůstek percentilů.** Pro všechny percentilové grafy toto nastavení řídí přírůstky v percentilech, které se zobrazují v grafu: 1, 2, 5, 10, 20 nebo 25.

Výběr výstupu grafu

1. V hlavním dialogovém okně Rozhodovací strom klepněte na volbu **Výstup**.
2. Klepněte na kartu **Ploty** .

## Výběr a pravidla přidělení skóre

Karta Pravidla poskytuje schopnost generovat pravidla výběru nebo klasifikace/predikce ve formě syntaxe příkazů, SQL nebo jednoduchých (plain-anglických) textu. Tato pravidla můžete zobrazit v prohlížeči a/ nebo uložit pravidla do externího souboru.

**Syntaxe.** Řídí formu pravidel výběru v obou výstupu zobrazených v Prohlížeči a pravidel výběru uložených do externího souboru.

- **IBM SPSS Statistics.** Jazyk syntaxe příkazu. Pravidla jsou vyjádřena jako sada příkazů, které definují podmínku filtru, kterou lze použít k výběru podmnožiny případů nebo jako COMPUTE příkazů, které lze použít k skóre případů.
- **SQL.** Jsou generována standardní pravidla SQL pro výběr nebo extrakci záznamů z databáze nebo pro přiřazení hodnot k těmto záznamům. Generovaná pravidla SQL nezahrnují žádná jména tabulek nebo jiné informace o zdroji dat.
- **Jednoduchý text.** Prostý anglický pseudo-kód. Pravidla jsou vyjádřena jako sada logických příkazů "if ... then", které popisují klasifikace nebo předpovědi modelu pro každý uzel. Pravidla v tomto formuláři mohou používat definovaná jména proměnných a hodnot nebo názvy proměnných a datové hodnoty.

**Typ.** Pro produkt IBM SPSS Statistics a pravidla SQL řídí typ generovaných pravidel: pravidla pro výběr nebo přidělení skóre.

- **Přiřad'te hodnoty k případům.** Pravidla lze použít k přiřazení předpovědi modelu k případům, které splňují kritéria členství uzlů. Pro každý uzel, který splňuje kritéria členství uzlu, se vygeneruje samostatné pravidlo.
- **Vyberte případy.** Pravidla lze použít k výběru případů, které splňují kritéria členství uzlů. V případě produktu IBM SPSS Statistics a pravidel SQL se vygeneruje jediné pravidlo pro výběr všech případů, které splňují kritéria výběru.

**Zahrnout náhradníky v produktu IBM SPSS Statistics a pravidel SQL.** Pro CRT a QUEST můžete zahrnout náhradní prediktory z modelu v pravidlech. Pravidla, která obsahují surrogates může být poměrně složitá. Obecně platí, že pokud chcete odvodit konceptuální informace o stromu, vylučte náhradní náhrady. Pokud některé případy mají nekompletní data nezávislé proměnné (prediktory) a chcete pravidla, která napodobují váš strom, obsahují náhradní řešení. Další informace naleznete v tématu [“Náhradníci”](#) na stránce 8 .

**Uzly.** Řídí rozsah generovaných pravidel. Pro každý uzel zahrnutý v rozsahu se vygeneruje samostatné pravidlo.

- **Všechny uzly terminálu.** Generuje pravidla pro každý uzel terminálu.
- **Nejlepší uzly terminálu.** Generuje pravidla pro prvních  $n$  uzlů terminálu založených na hodnotách indexu. Pokud číslo překročí počet uzlů terminálu ve stromu, pravidla se vygenerují pro všechny uzly terminálu. (Viz poznámka níže.)
- **Nejlepší uzly terminálu až do zadaného procenta případů.** Generuje pravidla pro uzly terminálu pro prvních  $n$  procent případů na základě hodnot indexu. (Viz poznámka níže.)
- **Uzly terminálu, jejichž hodnota indexu odpovídá nebo přesahuje hodnotu uzavření objektu.** Generuje pravidla pro všechny uzly terminálu s hodnotou indexu větší než nebo rovnou zadané hodnotě. Hodnota indexu větší než 100 znamená, že procentní část případů v cílové kategorii v daném uzlu překračuje procentní část v kořenovém uzlu. (Viz poznámka níže.)
- **Všechny uzly.** Generuje pravidla pro všechny uzly.

*Poznámka 1:* Výběr uzlu založený na hodnotách indexu je dostupný pouze pro kategoriální závislé proměnné s definovanými cílovými kategoriemi. Pokud jste zadali více cílových kategorií, je pro každou cílovou kategorii vygenerována samostatná sada pravidel.

*Poznámka 2:* Pro produkt IBM SPSS Statistics a pravidla SQL pro výběr případů (ne pro přiřazování hodnot), **Všechny uzly** a **Všechny uzly terminálu** budou efektivně generovat pravidlo, které vybere všechny případy použité v analýze.

**Exportovat pravidla do souboru.** Uloží pravidla do externího textového souboru.

Můžete také generovat a uložit pravidla výběru nebo přidělení pravidel interaktivně, na základě vybraných uzlů v konečném modelu stromu. Další informace naleznete v tématu [“Výběr případu a pravidla přidělení skóre”](#) na stránce 18 .

*Poznámka:* Pokud použijete pravidla ve formě syntaxe příkazu na jiný datový soubor, tento datový soubor musí obsahovat proměnné se stejnými názvy jako nezávislé proměnné zahrnuté v konečném modelu, měřeno ve stejné metrice, se stejnými uživatelem definovanými chybějícími hodnotami (pokud nějaké existují).

Uvést pravidla pro výběr nebo přidělení skóre

1. V hlavním dialogovém okně Rozhodovací strom klepněte na volbu **Výstup**.
2. Klepněte na kartu **Pravidla** .

## Editor stromu

---

Pomocí stromu Editor můžete:

- Skrýt a zobrazit vybrané větve stromu.
- Řízení zobrazení obsahu uzlu, statistiky zobrazované na rozdělení uzlu a další informace.
- Změnit barvy uzlu, pozadí, ohraničení, grafu a písma.
- Změnit styl a velikost písma.
- Změnit zarovnání stromu.
- Vyberte dílčí sady případů pro další analýzu založenou na vybraných uzlech.
- Vytvořit a uložit pravidla pro výběr nebo přidělení případů na základě vybraných uzlů.

Chcete-li upravit stromový model:

1. Poklepejte na stromový model v okně Prohlížeče.  
nebo
2. Z nabídky Upravit nebo z rozevírací nabídky klepněte pravým tlačítkem myši vyberte:

**Upravit**

Skrytí a zobrazení uzlů

Chcete-li skrýt (sbalit) všechny podřízené uzly ve větvi pod nadřazeným uzlem, postupujte takto:

1. Klepněte na znaménko minus (-) v malém poli pod dolním pravým rohem nadřazeného uzlu.

Všechny uzly pod nadřazeným uzlem v této větvi budou skryty.

Chcete-li zobrazit (rozbalit) podřízené uzly ve větvi pod nadřazeným uzlem, postupujte takto:

2. Klepněte na znaménko plus (+) v malém poli pod dolním pravým rohem nadřazeného uzlu.

*Poznámka:* Skrytí podřízených uzlů na větvi není stejné jako smazání stromu. Pokud chcete vyčištěný strom, musíte požadovat smazání dřívě, než vytvoříte strom, a vyčištěné větve nejsou zahrnuty do finálního stromu. Další informace naleznete v tématu [“Stromy stromů”](#) na stránce 7 .

Výběr více uzlů

Můžete vybrat případy, generovat pravidla pro přidělení skóre a výběry a provádět další akce založené na aktuálně vybraných uzlech. Výběr více uzlů:

1. Klepněte na uzel, který chcete vybrat.
2. Klepněte pravým tlačítkem myši na další uzly, které chcete vybrat.

Můžete vybrat více sourozeneckých uzlů a/nebo nadřazených uzlů v jedné větvi a podřízených uzlech v jiné větvi. Nemůžete však použít vícenásobný výběr v nadřazeném uzlu a podřízeným/následníka stejné větve uzlu.

## Práce s velkými stromy

Stromové modely mohou někdy obsahovat tolik uzlů a větví, že je obtížné nebo nemožné zobrazit celý strom v plné velikosti. Při práci s velkými stromy může být užitečné použít mnoho funkcí:

- **Stromová mapa.** Můžete použít stromovou mapu, mnohem menší, zjednodušenou verzi stromu, abyste mohli procházet stromem a vybírat uzly. Další informace naleznete v tématu [“Stromová mapa”](#) na stránce 16 .
- **Rozšiřitelnost.** Změna měřítka v procentech pro zobrazení stromu můžete zmenšit a zvětšit. Další informace naleznete v tématu [“Rozšiřitelnost zobrazení stromu”](#) na stránce 17 .
- **Zobrazení uzlu a větve.** Strom můžete vytvořit kompaktnější zobrazením pouze tabulek nebo pouze grafů v uzlech a/nebo potlačením zobrazení štítků uzlů nebo informací o independent proměnných. Další informace naleznete v tématu [“Řízení informací zobrazených ve stromu”](#) na stránce 17 .

## Stromová mapa

Stromová mapa poskytuje kompaktní, zjednodušené zobrazení stromu, které můžete použít k navigaci ve stromu a k výběru uzlů.

Chcete-li použít okno stromové mapy:

1. V nabídce Editor stromu zvolte:

**Zobrazit > Stromová mapa**

- Aktuálně vybraný uzel je zvýrazněn jak v editoru modelů stromu, tak v okně stromové mapy.
- Část stromu, která se aktuálně nachází v oblasti pohledu Editor modelů stromu, je v mapě stromu označena červeným obdélníkem. Chcete-li změnit sekci stromu zobrazeného v oblasti zobrazení, klepněte pravým tlačítkem myši a tažením označte obdélník.
- Vyberete-li uzel ve stromové mapě, který se aktuálně nachází v oblasti pohledu Editor stromu, zobrazí se změny, aby zahrnoval vybraný uzel.
- Vícenásobný výběr uzlu funguje stejně jako v mapě stromu jako v editoru stromu: Ctrl-click pro výběr více uzlů. Na nadřazeném uzlu nemůžete použít vícenásobný výběr a podřízený/podřízený prvek téže větve uzlu.

## Rozšiřitelnost zobrazení stromu

Ve výchozím nastavení se stromy automaticky přizpůsobují tak, aby se vešly do okna prohlížeče, což může mít za následek to, že některé stromy, které se zpočátku velmi obtížně čtou. Můžete vybrat přednastavené měřítko nebo zadat vlastní hodnotu vlastního měřítka mezi 5% a 200%.

Chcete-li změnit měřítko stromu:

1. Vyberte měřítko procentní části z rozevíracího seznamu na panelu nástrojů nebo zadejte vlastní procentní hodnotu.

nebo

2. V nabídce Editor stromu zvolte:

**Pohled > Měřítko ...**

Před vytvořením modelu stromu můžete také zadat hodnotu měřítka. Další informace naleznete v tématu [“Výstup” na stránce 11](#).

## Okno Souhrn uzlu

Souhrnné okno uzlu poskytuje větší zobrazení vybraných uzlů. Také můžete použít souhrnné okno k zobrazení, použití nebo uložení pravidel výběru nebo přidělení skóre na základě vybraných uzlů.

- Použijte nabídku Zobrazení v okně souhrnu uzlu pro přepnutí mezi zobrazeními souhrnné tabulky, grafu nebo pravidel.
- Použijte nabídku Pravidla v okně se souhrnem uzlu a vyberte typ pravidel, která chcete zobrazit. Další informace naleznete v tématu [“Výběr případu a pravidla přidělení skóre” na stránce 18](#).
- Všechny pohledy v okně Souhrn uzlu reflektují kombinovaný souhrn pro všechny vybrané uzly.

Chcete-li použít okno souhrnu uzlu:

1. Vyberte uzly v editoru stromu. Chcete-li vybrat více uzlů, klepněte na tlačítko Ctrl-click.
2. Z nabídky vyberte:

**Zobrazit > Souhrn**

## Řízení informací zobrazených ve stromu

Nabídka Volby v Editoru stromu vám umožňuje řídit zobrazení obsahu uzlu, názvů nezávislých proměnných (prediktorů) a statistiky, definic uzlů a dalších nastavení. Mnohé z těchto nastavení lze také ovládat z panelu nástrojů.

## Změna barev stromů a textových písem

Ve stromu lze změnit následující barvy:

- Okraj uzlu, pozadí a barva textu
- Barva větve a barva textu větve
- Barva pozadí stromu
- Barva zvýraznění zvýrazněné kategorie (kategorické závislé proměnné)
- Barvy grafu uzlů

Také můžete změnit písmo, styl a velikost písma pro veškerý text ve stromu.

*Poznámka:* Atributy barvy nebo písma nemůžete změnit pro jednotlivé uzly nebo větve. Změny barvy se vztahují na všechny prvky stejného typu a změny písma (jiné než barevné) se vztahují na všechny prvky grafu.

Chcete-li změnit barvy a atributy písma textu:

1. Použijte panel nástrojů ke změně atributů písma pro celý strom nebo barvy pro různé prvky stromu. (Popisy tlačítek popisují každý ovládací prvek na panelu nástrojů, když umístíte kurzor myši na ovládací prvek.)  
nebo
2. Poklepejte kdekoli v Editoru stromu a otevřete okno Vlastnosti nebo z nabídek vyberte:  
**Pohled > Vlastnosti**
3. U ohraničení, větve, pozadí uzlu, předpokládané kategorie a stromového pozadí klepněte na kartu **Barva**.
4. V případě barev a atributů písma klepněte na kartu **Text**.
5. Chcete-li barvy grafu uzlu, klepněte na kartu **Grafy uzlů**.

## Výběr případu a pravidla přidělení skóre

Editor stromu můžete použít k:

- Vyberte podmnožiny případů na základě vybraného uzlu (ů). Další informace naleznete v tématu [“Filtrování případů” na stránce 18](#).
- Generovat pravidla pro výběr případu nebo pravidla přidělení pravidel v syntaxi příkazů produktu IBM SPSS Statistics nebo ve formátu SQL. Další informace naleznete v tématu [“Ukládání pravidel pro výběr a přidělení skóre” na stránce 18](#).

Můžete také automaticky ukládat pravidla na základě různých kritérií, když spustíte proceduru Rozhodovací strom pro vytvoření stromového modelu. Další informace naleznete v tématu [“Výběr a pravidla přidělení skóre” na stránce 14](#).

### Filtrování případů

Chcete-li vědět více o případu v konkrétním uzlu nebo skupině uzlů, můžete vybrat podmnožinu případů pro další analýzu založenou na vybraných uzlech.

1. Vyberte uzly v editoru stromu. Chcete-li vybrat více uzlů, klepněte na tlačítko Ctrl-click.
2. Z nabídky vyberte:  
**Pravidla > Filtrovat případy ...**
3. Zadejte název proměnné filtru. Případy z vybraných uzlů obdrží hodnotu 1 pro tuto proměnnou. Všechny ostatní případy budou mít hodnotu 0 a budou vyloučeny z následné analýzy, dokud nezměníte stav filtru.
4. Klepněte na tlačítko **OK**.

### Ukládání pravidel pro výběr a přidělení skóre

Můžete uložit výběr případu nebo pravidla přidělení pravidel v externím souboru a pak použít tato pravidla na jiný zdroj dat. Pravidla jsou založena na vybraných uzlech v Editoru stromu.

**Syntaxe.** Řídí formu pravidel výběru v obou výstupu zobrazených v Prohlížeči a pravidel výběru uložených do externího souboru.

- **IBM SPSS Statistics.** Jazyk syntaxe příkazu. Pravidla jsou vyjádřena jako sada příkazů, které definují podmínku filtru, kterou lze použít k výběru podmnožiny případů nebo jako COMPUTE příkazů, které lze použít k skóre případů.
- **SQL.** Jsou generována standardní pravidla SQL k výběru/extrahování záznamů z databáze nebo k přiřazení hodnot k těmto záznamům. Generovaná pravidla SQL nezahrnují žádná jména tabulek nebo jiné informace o zdroji dat.

**Typ.** Můžete vytvořit pravidla pro výběr nebo přidělení skóre.

- **Vyberte případy.** Pravidla lze použít k výběru případů, které splňují kritéria členství uzlů. V případě produktu IBM SPSS Statistics a pravidel SQL se vygeneruje jediné pravidlo pro výběr všech případů, které splňují kritéria výběru.
- **Přiřadte hodnoty k případům.** Pravidla lze použít k přiřazení předpovědí modelu k případům, které splňují kritéria členství uzlů. Pro každý uzel, který splňuje kritéria členství uzlu, se vygeneruje samostatné pravidlo.

**Zahrnout náhradníky.** Pro CRT a QUEST můžete zahrnout náhradní prediktory z modelu v pravidlech. Pravidla, která obsahují surrogates může být poměrně složitá. Obecně platí, že pokud chcete odvodit konceptuální informace o stromu, vylučte náhradní náhrady. Pokud některé případy mají nekompletní data nezávislé proměnné (prediktory) a chcete pravidla, která napodobují váš strom, obsahují náhradní řešení. Další informace naleznete v tématu [“Náhradníci”](#) na stránce 8 .

Chcete-li uložit pravidla pro výběr případu nebo přidělení skóre:

1. Vyberte uzly v editoru stromu. Chcete-li vybrat více uzlů, klepněte na tlačítko Ctrl-click.
2. Z nabídky vyberte:

**Pravidla > Exportovat ...**

3. Vyberte typ pravidel, které chcete, a zadejte název souboru.

*Poznámka:* Pokud použijete pravidla ve formě syntaxe příkazu na jiný datový soubor, tento datový soubor musí obsahovat proměnné se stejnými názvy jako nezávislé proměnné zahrnuté v konečném modelu, měřeno ve stejné metrice, se stejnými uživatelem definovanými chybějícími hodnotami (pokud nějaké existují).





## Upozornění

---

Tyto informace byly vytvořeny pro produkty a služby poskytované v USA. Tento materiál může být dostupný od IBM v jiných jazycích. K povolení přístupu však může být vyžadováno vlastnictví kopie produktu nebo verze produktu v tomto jazyce.

Společnost IBM nemusí nabízet produkty, služby nebo funkce uvedené v tomto dokumentu v jiných zemích. Informace o produktech a službách, které jsou aktuálně k dispozici ve vaší oblasti, získáte od lokálního zástupce společnosti IBM. Odkazy na produkty, programy nebo služby společnosti IBM neuvádí ani neimplikují, že lze použít pouze daný produkt, program nebo službu společnosti IBM. Lze použít libovolný funkčně ekvivalentní produkt, program nebo službu neporušující práva duševního vlastnictví společnosti IBM. Vyhodnocení a ověření funkčnosti produktů, programů nebo služeb, které nepatří společnosti IBM, je však zodpovědností uživatele.

Společnost IBM může vlastnit patenty nebo nevyřízené žádosti o patenty zahrnující předměty popsané v tomto dokumentu. Vlastnictví tohoto dokumentu neposkytuje licenci k těmto patentům. Dotazy na licence můžete písemně odeslat na následující adresu:

*IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
USA*

S dotazy na licence týkající se dvoubajtových informací (DBCS) se obraťte na oddělení intelektuálního vlastnictví společnosti IBM v dané zemi, nebo je odešlete písemně na následující adresu:

*Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.  
19-21, Nihonbashi-Hakozakicho, Chuo-ku  
Tokyo 103-8510, Japan*

SPOLEČNOST INTERNATIONAL BUSINESS MACHINES CORPORATION POSKYTUJE TUTO PUBLIKACI "TAKOVOU, JAKÁ JE", BEZ JAKÝCHKOLIV ZÁRUK, VYJÁDŘENÝCH NEBO ODVOZENÝCH VČETNĚ, MIMO JINÉ, ODVOZENÝCH ZÁRUK NEPORUŠENÍ PRÁV TŘETÍCH STRAN, ZÁRUKY PRODEJNOSTI NEBO VHODNOSTI PRO URČITÝ ÚČEL. Některé právní řády u určitých transakcí nepřipouštějí vyloučení záruk výslovně vyjádřených nebo vyplývajících z okolností, a proto se na vás výše uvedené omezení nemusí vztahovat, a proto se vás toto prohlášení nemusí týkat.

Uvedené údaje mohou obsahovat technické nepřesnosti nebo typografické chyby. Údaje zde uvedené jsou pravidelně upravovány a tyto změny budou zahrnuty v nových vydáních této publikace. Společnost IBM může kdykoli bez upozornění provádět vylepšení nebo změny v produktech či programech popsaných v této publikaci.

Jakékoliv odkazy v této publikaci na webové stránky jiných společností než IBM jsou poskytovány pouze pro pohodlí uživatele a nemohou být žádným způsobem vykládány jako doporučení těchto webových stránek. Materiály uvedené na těchto webových stránkách nejsou součástí materiálů pro tento produkt IBM a použití uvedených stránek je pouze na vlastní nebezpečí.

IBM může použít nebo distribuovat jakékoli informace, které jí poskytnete, libovolným způsobem, který společnost považuje za odpovídající, bez vzniku jakýchkoliv závazků vůči vám.

Vlastníci licence k tomuto programu, kteří chtějí získat informace o možnostech (i) výměny informací s nezávisle vytvořenými programy a jinými programy (včetně tohoto) a (ii) oboustranného využití vyměňovaných informací, mohou kontaktovat informační středisko na adrese:

*IBM Director of Licensing  
IBM Corporation*

North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
USA

Poskytnutí takových informací může být podmíněno dodržením určitých podmínek a požadavků zahrnujících v některých případech uhrazení stanoveného poplatku.

Licencovaný program popsáný v tomto dokumentu a veškerý licencovaný materiál k němu dostupný jsou společností IBM poskytovány na základě podmínek uvedených ve smlouvách IBM Customer Agreement, IBM International Program License Agreement nebo v jiné ekvivalentní smlouvě.

Citovaná data o výkonu a příklady klienta jsou uvedeny pouze pro názornost. Skutečné výsledky výkonu se mohou lišit v závislosti na specifických konfiguracích a provozních podmínkách.

Informace týkající se produktů jiných společností než IBM byly získány od dodavatelů těchto produktů, z jejich publikovaných sdělení, nebo z jiných veřejně dostupných zdrojů. IBM tyto produkty netestovala a nemůže potvrdit přesnost údajů o výkonu, kompatibilitě nebo jiná tvrzení týkající se produktů jiných společností než IBM. Otázky týkající se možností produktů jiných společností než IBM by měly být adresovány dodavatelům těchto produktů.

Prohlášení týkající se budoucího směru vývoje nebo záměrů společnosti IBM se mohou změnit nebo mohou být zrušena bez předchozího upozornění a představují pouze cíle a záměry.

Tyto údaje obsahují příklady dat a sestav používaných v běžných obchodních operacích. Aby byla představa úplná, používají se v příkladech jména osob, společností, značek a produktů. Všechna tato jména jsou fiktivní a jakákoliv podobnost se skutečnými lidmi nebo obchodními podniky je čistě náhodná.

#### COPYRIGHT - LICENCE:

Tyto informace obsahují ukázkové aplikační programy ve zdrojovém jazyku a ilustrují různé programovací techniky na různých operačních platformách. Tyto ukázkové programy můžete bez závazků vůči společnosti IBM jakýmkoli způsobem kopírovat, měnit a distribuovat za účelem vývoje, používání, odbytu či distribuce aplikačních programů odpovídajících rozhraní API pro operační platformu, pro kterou byly ukázkové programy napsány. Tyto příklady nebyly důkladně testovány ve všech podmínkách. Společnost IBM proto nemůže zaručit spolehlivost, upotřebitelnost nebo funkčnost těchto programů. Ukázkové programy jsou poskytovány "JAK JSOU", bez záruky jakéhokoli druhu. IBM nenes odpovědnost za žádné škody vzniklé ve spojení s Vaším užíváním ukázkových programů.

Jakákoli kopie nebo část těchto ukázkových programů nebo jakékoli odvozené dílo musí obsahovat následující poznámku o autorských právech:

© Copyright IBM Corp. 2021. Části tohoto kódu jsou odvozeny ze vzorových programů společnosti IBM Corp. Vzorové programy.

© Copyright IBM Corp. 1989-2021. Všechna práva vyhrazena.

## Ochranné známky

---

IBM, logo IBM a [ibm.com](http://ibm.com) jsou ochranné známky nebo registrované ochranné známky společnosti International Business Machines Corp., registrované v mnoha jurisdikcích po celém světě. Ostatní názvy produktů a služeb mohou být ochrannými známkami společnosti IBM nebo jiných společností. Aktuální seznam ochranných známek společnosti IBM je k dispozici na webu na stránce "Copyright and trademark information" na adrese [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Adobe, logo Adobe, PostScript a logo PostScript jsou buď registrované ochranné známky, nebo ochranné známky společnosti Adobe Systems Incorporated ve Spojených státech anebo v dalších zemích.

Intel, logo Intel, Intel Inside, logo Intel Inside, Intel Centrino, logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium a Pentium jsou ochranné známky nebo registrované ochranné známky společnosti Intel Corporation nebo jejich dceřiných společností ve Spojených státech a případně v dalších jiných zemích.

Linux je registrovaná ochranná známka Linuse Torvaldse ve Spojených státech a případně v dalších jiných zemích.

Microsoft, Windows, Windows NT a logo Windows jsou ochranné známky společnosti Microsoft Corporation ve Spojených státech a případně v dalších jiných zemích.

UNIX je registrovaná ochranná známka společnosti The Open Group ve Spojených státech a případně v dalších jiných zemích.

Java a všechny ochranné známky a loga založené na jazyce Java jsou ochranné známky nebo registrované ochranné známky společnosti Oracle anebo příbuzných společností.



# Rejstřík

## C

### CRT

Nečistota [6](#)  
smazání [7](#)

## Č

### Číslo uzlu

uložení jako proměnné z rozhodovacích stromů [11](#)

## D

dvouvčí [6](#)

## G

Gini [6](#)

## H

hodnoty indexu  
stromy [12](#)

## CH

chybějící hodnoty  
stromy [10](#)  
chybná klasifikace  
Náklady [8](#)  
stromy [12](#)

## K

křížovou validaci  
stromy [4](#)

## M

mazání rozhodovacích stromů  
a skrývání uzlů [7](#)

## N

Náklady  
chybná klasifikace [8](#)  
nečistota  
CRT stromy [6](#)

## O

objednané dvě [6](#)  
odhady rizik  
stromy [12](#)  
ověření

ověření (*pokračování*)  
stromy [4](#)

## P

### pravidla

vytvoření syntaxe výběru a skóre pro rozhodovací stromy  
[14](#), [18](#)

### predikovaná pravděpodobnost

uložení jako proměnné z rozhodovacích stromů [11](#)

### predikované hodnoty

uložení jako proměnné z rozhodovacích stromů [11](#)

## Q

### QUEST

smazání [7](#)

## R

rozdělení-ukázka ověření  
stromy [4](#)

### rozhodovací stromy

metoda CRT [1](#)

Metoda CHAID [1](#)

Metoda QUEST [1](#), [7](#)

úroveň měření [1](#)

Vyčerpávající metoda CHAID [1](#)

vynucení první proměnné do modelu [1](#)

## Ř

### ŘETĚŽIKO

intervaly pro měřítko nezávislých proměnných [6](#)

maximální iterace [5](#)

opětovné splezení sloučených kategorií [5](#)

rozdělení a sloučení kritérií [5](#)

Úprava Bonferroniho [5](#)

## S

sbalení větví stromů [15](#)

semeno náhodných čísel

ověření stromu rozhodnutí [4](#)

skóre

stromy [10](#)

skrytí uzlů

versus promazání [7](#)

skrytí větví stromu [15](#)

SQL

vytvoření syntaxe SQL pro výběr a hodnocení [14](#), [18](#)

stromy

barvy [17](#)

barvy diagramu uzlu [17](#)

důležitost prediktoru [12](#)

generování pravidel [14](#), [18](#)

## stromy (pokračování)

- grafy [13](#)
- hodnoty indexu [12](#)
- chybějící hodnoty [10](#)
- intervaly pro měřítko nezávislých proměnných [6](#)
- Kritéria pro pěstování CHAID [5](#)
- křížovou validaci [4](#)
- metoda CRT [6](#)
- náklady na nesprávné klasifikace [8](#)
- obsah stromu v tabulce [12](#)
- odhady rizik [12](#)
- omezení počtu úrovní [5](#)
- orientace stromu [12](#)
- písmata [17](#)
- práce s velkými stromy [16](#)
- předchozí pravděpodobnost [9](#)
- rozdělení-ukázka ověření [4](#)
- řízení velikosti uzlu [5](#)
- skóre [10](#)
- skrytí větví a uzlů [15](#)
- smazání [7](#)
- statistika uzlu terminálu [12](#)
- stromová mapa [16](#)
- tabulka misclassification [12](#)
- textové atributy [17](#)
- ukládání proměnných modelu [11](#)
- úpravy [15](#)
- výběr více uzlů [15](#)
- Zisk [9](#)
- zobrazení a skrytí statistiky větví [12](#)
- zobrazení řídicího stromu [12](#), [17](#)
- zobrazení stromu škálování [17](#)

## syntaxe

- vytvoření syntaxe výběru a skóre pro rozhodovací stromy [14](#), [18](#)

## Syntaxe příkazu

- vytvoření syntaxe výběru a skóre pro rozhodovací stromy [14](#), [18](#)

## U

### úroveň měření

- rozhodovací stromy [1](#)

### úroveň významnosti pro rozdělování uzlů [7](#)

### uzly

- výběr více uzlů stromu [15](#)

## V

### váhové případy

- zlomkové váhy v rozhodovacích stromech [1](#)

- výběr více uzlů stromu [15](#)

## Z

### Zisk

- předchozí pravděpodobnost [9](#)
- stromy [9](#), [12](#)



