

IBM SPSS DECISION TREES

Jednoduše identifikujte skupiny a predikujte

Stromově uspořádané postupné štěpení dat na homogenní podmnožiny je technika vhodná pro exploraci vztahů i pro tvorbu rozhodovacích pravidel. Využívá se ve statistické analýze dat i v dataminingovém rozhodovacím procesu a skórování.

Modul IBM SPSS Decision Trees nabízí klasifikační, rozhodovací a asociační stromy pro identifikaci skupin, analýzu vztahů mezi skupinami i pro predikci hodnot. Klasifikační a rozhodovací stromy využijete pro segmentaci, stratifikaci, predikci i pro redukci dat a screening, identifikaci interakcí, slučování kategorií a diskretizaci spojitých proměnných. Asociační interpretace stromů uplatníte při tvorbě i ověřování hypotéz o kauzálních vazbách mezi proměnnými a při exploračním post hoc seskupení průměrů v jednoduché ANOVA či řádků v kontingenční tabulce.

Intuitivní a názorné zobrazení pomocí stromu dovoluje prezentovat vztahy jasným a přehledným způsobem i laickým uživatelům, což v praxi usnadňuje jejich rozhodování. Stromové grafy rovněž dovolují vizuálně posoudit výsledky a vhodnost modelu. Vizuální analýzou stromového větvení naleznete specifické podskupiny a vztahy, které nejsou na první pohled zřejmé nebo které by tradiční statistika dokázala odhalit pouze obtížně a zdlouhavě. Strom také odstupňuje sílu vlivu na cílovou proměnnou. Klasifikační postupy, a tím i pravidla, jsou v grafu zobrazeny jako větvení a uzly reprezentují podskupiny. Klasifikační pravidlo vás tímto způsobem postupně vede k přesnějšímu určení cílové proměnné a finálních tříd klasifikace a predikce; vizualizace hierarchickým štěpením zcela jasně ilustruje postupné kroky a též i statistický popis všech dílčích i konečných výsledků.

V grafické reprezentaci rozhodovacího procesu je zřetelně vidět, jaké skupiny vznikají štěpením podle daného pravidla a u kterých skupin pravidlo končí. Pořadí vstupu proměnných do stromu v procesu jeho růstu představuje indikaci síly vlivu i predikci pro jednotlivé proměnné a interakcí mezi nimi.

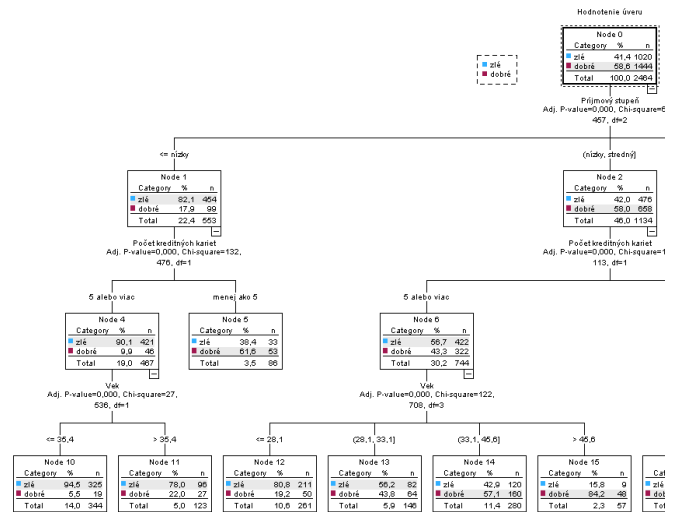
Nejdůležitější:

IBM SPSS Decision Trees umožňuje:

- identifikovat v datech vzory, segmenty a skupiny pomocí přehledné vizualizace
- zvolit ze čtyř zavedených algoritmů pro růst stromu: CHAID, Exhaustive CHAID, C&RT a QUEST
- prezentovat velmi přehledně a jednoduše výsledky i laikům
- na základě získaného klasifikačního stromu vytvořit nové proměnné pro případy v datové matici (koncový uzel, predikce a konfidence odhadu)

IBM SPSS Decision Trees použijete pro:

- *databázový marketing*
 - *výběr vhodných proměnných pro segmentaci databáze zákazníků (odpověděl/neodpověděl na testovací mail či na předchozí podobné kampaně, patří mezi top/průměrné/nevýznamné zákazníky, má/nemá rozšířené služby apod.)*
 - *profilování skupin zákazníků jinými atributy (např. měřitelnými či disponibilními), jako jsou demografické údaje a chování či aktivita zákazníka*
 - *příprava nové reklamní kampaně, cílená na specifické podskupiny vedoucí ke snížení nákladů a zlepšení ROI*
- *výzkum trhu*
 - *výzkum spokojenosti zákazníků a zaměstnanců*
 - *stanovení proměnných determinujících spokojenost (např. na škále 1 až 7)*
 - *profilování úrovně spokojenosti na základě odpovědi na vybrané sociodemografické, postojové charakteristiky a dílčí spokojenosti či otázky o potřebách respondentů*
 - *opatření vycházející z těch kombinací nalezených faktorů a vlivů, které vedou k vysoké spokojenosti, a tedy i k retenci a loajalitě*
- *kreditní skórování a skórování rizik*
 - *stanovení rizikových skupin (silně/středně/málo rizikové)*
 - *nalezení rizikových skupin podle informací o zákaznících, např. o jejich transakcích na účtech nebo sociodemografických a ekonomických charakteristikách*
 - *nabídka vhodných služeb pravé skupině žadatelů o úvěr, a to podle rizik, která jsou s ní spojena*
- *cílení programu*
 - *identifikace kritériální proměnné, která charakterizuje úspěšnost programu*
 - *nalezení faktorů vedoucích k úspěchu, díky informaci uložené v žádostech o program nebo z historických dat o předchozích programech*
 - *výběr a přizpůsobení nových programů pro uspokojení potřeb*
- *marketing ve státní sféře*
 - *výběr cílové proměnné pro segmentaci databáze zákazníků (např. potenciální žadatelé o vysokoškolské studium, kteří již žádali, proti těm, kteří ještě nežádali)*
 - *profilování skupin založené na predikčních attributech, jako např. demografických údajích či informacích o aktivitách zájemce*
 - *příprava nové reklamní kampaně, cílené na specifické podskupiny vedoucí ke snížení nákladů a zlepšení ROI*



Vizualizace stromů v IBM SPSS Decision Trees pro snadnou interpretaci výsledků a nalezení skrytých závislostí.

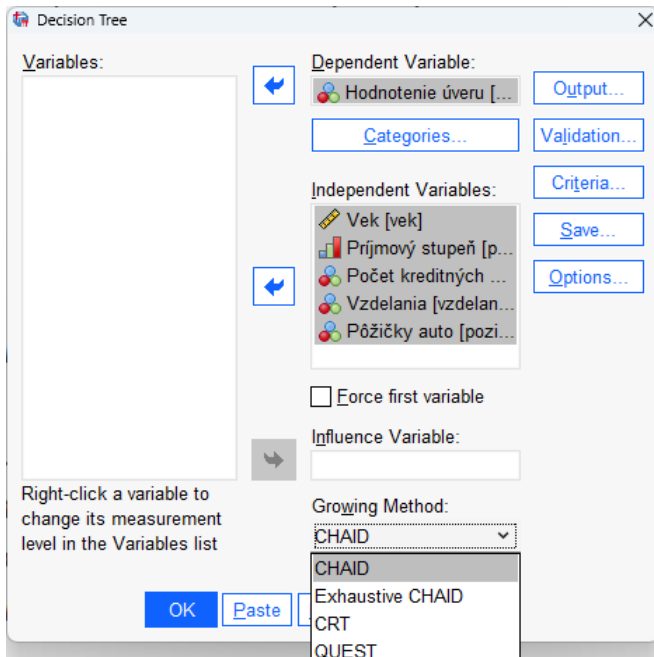
IBM SPSS Decision Trees je k dispozici pro instalaci v režimu lokální desktopové aplikace, avšak při požadavku na větší výkon a škálovatelnost lze provést také instalaci v režimu klient/server.

Vyberte si jeden ze čtyř algoritmů pro růst stromu

IBM SPSS Decision Trees obsahuje čtyři algoritmy pro růst stromu:

- **CHAID** - rychlý, statistický algoritmus založený na optimální hodnotě testu χ^2 nebo Ftestu; zkoumá data rychle a efektivně a tvoří segmenty a profily v sekvenci optimálních kroků; štěpí skupiny vždy na vhodný počet statisticky homogenních podskupin; je založen na statistickém testování hypotéz na každém kroku, optimální krok v pravidle je dán maximální signifikancí;
- **Exhaustive CHAID** - modifikace algoritmu CHAID, která prozkoumává všechna možná štěpení pro každý prediktor;
- **CRT (Classification & Regression Trees)** - algoritmus, který vytváří binární stromy (skupina je štěpena vždy na dvě části) a maximálně homogenní skupiny podle statistických kritérií;
- **QUEST** - statistický algoritmus, který vybírá proměnné bez vychýlení a vytváří přesné binární stromy rychle a efektivně

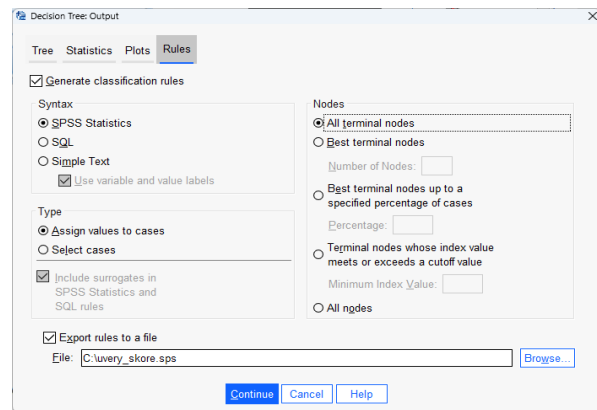
Tyto čtyři algoritmy nabízejí různé metody „růstu“ stromu, vyzkoušejte je a nalezněte tu nejlepší a nejvhodnější pro vaše data.



V IBM SPSS Decision Trees vytvoříte stromy pomocí metod CHAID, Exhaustive CHAID, C&RT a QUEST.

Rozšiřte své výsledky o další analýzy

Jestliže užíváte IBM SPSS Decision Trees společně s IBM SPSS Statistics Base, můžete při vytváření stromů snadno vložit výslednou klasifikační proměnnou do pracovního souboru a takto definovat skupiny přímo v datech nebo tuto proměnnou využít jako vstup pro další analýzy. Mimo to lze zapsat klasifikační/predikční pravidla pro jednotlivé skupiny ve formě syntaxe IBM SPSS Statistics, SQL dotazu či prostého textu, což otevírá možnost využití výsledků v budoucích aplikacích - při skórování, rozhodování, predikcích či při kontrolách asociací nezávisle na současném běhu a datech. Tato pravidla si prohlédnete ve výstupu a uložíte je do externího souboru pro pozdější predikce či zařazení nových záznamů. Jestliže chcete využít výsledky pro skórování jiných datových souborů, můžete zapsat informaci z tohoto modelu přímo do vašich dat nebo vytvořit XML model pro použití s IBM SPSS Statistics Server.



Export pravidel pro výběr případů a predikování k dalšímu využití vytvořeného modelu.

SPECIFIKACE

Stromy

- zobrazení stromových diagramů, map stromů, sloupcových grafů a četnostních tabulek
- jednoduchá tvorba stromu díky komplexnímu rozhraní, umožňujícímu tato nastavení:
 - typ proměnné (nominální, ordinální, číselná)
 - nezávislé proměnné
 - závislé proměnné
 - proměnná vlivu
 - algoritmus růstu stromu
 - výběr výstupů (strom, statistiky, grafy, pravidla)
 - design evaluace stromu
 - rozdělení souboru na trénovací a testovací množinu
 - křížová validace
 - zastavovací pravidla
 - uložení proměnných včetně predikovaných hodnot, pravděpodobností a modelu ve formátu XML
- čtyři algoritmy růstu stromu
- zobrazení rozložení cílové proměnné v uzlu
 - frekvenční tabulky
 - sloupcové grafy rozložení ve skupinách
- skrývání větví, změna vzhledu, fontů a barev
- prohlížení a tisk stromů
- přiblížení a oddálení stromu
- automatický růst stromu v produkčním módu
 - automatický zápis syntaxe z dialogového okna
- volba prvního prediktora
- orientační mapa stromu pro jeho prohlížení
- volba orientace grafu
 - horizontální pravolevá nebo levoprává
 - vertikální
- zadání apriorních pravděpodobností, penalizací chyb, nákladů, výnosů a vzdálenostních skóre kategorií závislé ordinální proměnné

Algoritmy růstu

- *analýza dat jedním ze čtyř algoritmů růstu:*
 - CHAID (Kass; 1980)
 - Exhaustive CHAID (Biggs, de Ville, Suen; 1991)
 - Classification & Regression Trees (CRT) (Breiman, Friedman, Olshen, Stone; 1984)
 - QUEST (Loh, Shih; 1997)
- *ošetření chybějících hodnot prediktorů jednou ze dvou metod: zavedení samostatné kategorie, nebo využití náhradního štěpení podle prediktoru s platnou hodnotou*
- *diskretizace spojitých proměnných na zvolený počet intervalů*
- *prořezávání stromů u algoritmů CRT a QUEST*
- *náhodné rozdělení souboru na tréninkovou a testovací množinu nebo rozdělení podle hodnot vybrané proměnné*

Evaluace kvality modelu

- *klasifikační tabulka a podíl chybných predikcí, standardní chyba odhadu, reziduální součet čtverců pro spojitou cílovou proměnnou*
- *evaluační grafy a tabulky pro identifikaci vhodných skupin (uzlů):*
 - *přínos / zisk (Gain chart)*
 - *index (Lift chart)*
 - *odezva (Response chart)*
 - *průměr (Mean chart)*
 - *očekávaný profit (Average Profit chart)*
 - *ROI (ROI chart)*
- *rozdělení dat na tréninkovou a testovací skupinu pro ověření přesnosti*
- *zobrazení grafů nebo klasifikačních pravidel pro vybrané uzly v samostatném okně*

Aplikování modelu

- *export diagramů, grafů a tabulek*
 - *formáty HTML, text, Word/RTF, Excel® a PDF*
- *uložení informací a predikcí modelu do proměnných v pracovním datovém souboru*
- *export rozhodovacích pravidel definujících uzly do SQL (skórování v databázi), do syntaxového jazyka IBM SPSS Statistics (skórování v systému IBM SPSS Statistics) nebo jako prostý text*
- *export modelů do XML pro skórování nových případů v aplikacích IBM SPSS Statistics Server a IBM SPSS Statistics Portal*
- *zveřejnění modelů jako obrázky nebo statické, či interaktivní tabulky pomocí IBM SPSS Statistics Portal*
- *výběr případů pro další analýzy v pracovním datovém souboru pomocí označených uzlů stromu*

IBM Business Analytics

IBM Business Analytics software poskytuje akční vhled, který manažeři potřebují k dosažení lepších obchodních výsledků. IBM nabízí komplexní, jednotné portfolio složené z business intelligence, prediktivních a sofistikovaných analýz, finančních výpočtů a strategického managementu, kontroly rizik a podmínek a analytických aplikací.

Pomocí IBM software dokážou společnosti odhalit trendy, vzory chování a odchylky od nich, porovnat scénáře „co by bylo kdyby“, predikovat potencionální hrozby a příležitosti, identifikovat a řídit klíčová obchodní rizika a plány, rozvrhnout a odhadovat zdroje. Díky těmto analytickým možnostem dokážou naši zákazníci po celém světě lépe pochopit, předvídat a formovat obchodní výsledky.