

IBM SPSS Modeler Personal

Kvalitnější rozhodování z prostředí Vašeho počítače

Data mining poskytuje organizacím jasnější pohled na současný stav a hlubší vhled do budoucnosti, založený na analýze historických dat. IBM SPSS Modeler je dataminingový nástroj pro analýzu dat a extrakci predikční znalosti. Tato znalost může pak být využita pro efektivní obchodní rozhodování.

Společnosti z různých odvětví používají SPSS Modeler a extrahované predikční znalosti pro přilákání nových a posílení loajality stávajících zákazníků. Mohou také snadněji a cenově efektivněji identifikovat zákazníky s potenciálem odchodu ke konkurenci (churn). Rovněž jsou schopni minimalizovat rizika a identifikovat podvodné chování (fraud). Organizacím z veřejného sektoru pomůže SPSS Modeler predikovat potřebu pracovních kapacit a proaktivně řešit otázky veřejné bezpečnosti. V akademické sféře může být SPSS Modeler použit při řízení institucí a při výuce.

Od verze 17 je tento nástroj k dispozici i v osobní (personal) edici.

Hlavní rysy

- podporuje přesnější predikce identifikací skrytých vzorů a trendů v datech
- díky uživatelsky intuitivnímu vizuálnímu zápisu posloupnosti datových manipulací nabízí jednoduchý přístup k datům, jejich přípravě a modelování
- poskytuje sofistikované technicky statistického i strojového učení pro rychlou tvorbu a validaci modelů
- rozšiřuje analytické nástroje o geoprostorové analýzy a umožňuje generování mapových výstupů v ESRI formátu

Predikční analýzy na vašem počítači

IBM SPSS Modeler Personal dovoluje vytvářet predikční modely, které umožňují objevit vztahy a trendy skryté v datech, z prostředí vašeho počítače. Obchodní problémy tak můžete vyřešit rychleji, pomocí silného data miningového nástroje s ověřenými analytickými technikami. Navíc SPSS Modeler automaticky připraví, vyčistí a transformuje data a zajistí vám tak kvalitní data pro vaši práci.

Analytik i běžný uživatel mohou s SPSS Modeler:

- získat, na datech založený, vzhled do obchodních procesů a realizovat jejich zlepšení
- pracovat s různými datovými formáty na vstupu i výstupu
- připojit se k podnikovým databázím
- vytvořit modely na strukturovaných datech
- optimalizovat vybrané modely podložené statistickými mírami spolehlivosti
- spravovat celý dataminingový cyklus efektivně z prostředí jednoho nástroje

Přímočarý dataminingový proces na klientském počítači

Intuitivní grafické rozhraní nástroje SPSS Modeler dovoluje uživatelům vizualizovat každý krok dataminingového procesu. Tuto vizualizaci označujeme jako proud (stream). Díky proudům mohou analytici i běžní uživatelé jednoduše spolupracovat a vnést obchodní znalost do dataminingových procesů. Uživatelé se více zaměří na objevování nové znalosti než na samotné technické řešení úlohy nebo dokonce psaní programového kódu. Uživatelé mohou volně a hlouběji prozkoumávat svá data a objevovat v nich další smysluplné vztahy.

Automatizovaná příprava dat a modelování dovoluje i těm, kteří nejsou analytici, vytvářet přesné modely rychle a jednoduše. Profesionální analytici mohou s výhodou využít pokročilých funkcionalit datové přípravy a predikčního modelování pro vytvoření ještě sofistikovanějších proudů.

Z vizuálního rozhraní SPSS Modeler Personal lze jednoduše přistoupit k jakémukoli datovému zdroji, ať už se jedná o databáze, textové soubory nebo soubory formátu IBM SPSS Statistics, IBM SPSS Data Collection, IBM Cognos

Business Intelligence, SAS nebo Microsoft Excel. K dispozici jsou tutoriály a integrovaná nápověda.

Vlastnosti a přínosy

- intuitivní rozhraní dovoluje různě zkušeným uživatelům vytvářet predikční modely
- automatizace přípravy dat umožní data rychle vyčistit
- nástroje na čištění dat nabízí odstraňování nebo nahrazování chybných údajů, automatické nahrazení chybějících hodnot a úpravy odlehklých a extrémních hodnot
- automatizované modelování zajistí zpřesnění vašich analýz
- jazyk R je nyní hlouběji integrován s nástrojem SPSS Modeler

Bohatá nabídka modelovacích technik a algoritmů

SPSS Modeler Personal nabízí pokročilé dataminingové techniky potřebné ve všech dataminingových úlohách:

- **klasifikační a regresní algoritmy.** Vytvářejí predikce nebo předpovědi založené na historických datech pomocí technik jako jsou rozhodovací stromy, neuronové sítě, logistická regrese, Coxova regrese, zobecněné lineární smíšené modely (GLMM) a další. Automatizované modelování je k dispozici pro binární i numerické cílové proměnné. Automatizovanou klasifikaci lze rovněž využít při aplikaci více algoritmů v jednom kroku. Nebudete se tak muset složitě rozhodovat, kdy použít správnou techniku
- **segmentační algoritmy.** Používají se pro seskupení objektů nebo nalezení nezvyklých vzorů v datech pomocí automatizovaného seskupování, detekce anomálií a klastrovacích technik neuronových sítí. Lze využít i automatizované seskupování
- **asociační algoritmy.** Objevte asociace, spojení a návaznosti pomocí algoritmů Apriori, CARMA. Můžete zkoumat i sekvenční asociace
- **časové řady a předpovědi.** Využijte statistických modelovacích technik pro tvorbu předpovědí jedné či více časových řad

Geoprostorová data

Prostorové analýzy v SPSS Modeler Personal umožňují objevovat vztahy v datech s ohledem na lokalitu jejich výskytu. Geoprostorová analýza dat může odhalit souvislosti, které by z grafů nebo tabulek nebyly patrné. Pro zobrazení výsledků geoprostorové analýzy se používají mapové podklady formátu ESRI. Analýzou neprostorových i prostorových dat můžete zlepšit celkovou přesnost modelu.

Do svých analýz můžete přidat úplně novou dimenzi tím, že objevíte vztahy mezi prostorovými a neprostorovými atributy. Časoprostorové predikce mohou výrazně vylepšit lineární modely přidáním informace o umístění ve dvourozměrném prostoru. To může přispět k identifikaci aktivních míst a predikci jejich vývoje v čase. Geoprostorová data mining najde uplatnění například v analýze kriminality, predikci vývoje epidemií, při správě budov nebo analýze výkonnosti poboček (branch performance analysis).

Integrace jazyka R

Společným užitím nástroje SPSS Modeler a programovacího jazyka R získáte z obou maximum. Z rozhraní SPSS Modeler Personal lze přímo spouštět syntaxi jazyka R. Jednoduše tak rozšíříte nativní možnosti datových manipulací a statistických a predikčních algoritmů o vlastní R kód. Můžete provádět analýzy, pracovat s výstupem a integrovat funkcionalitu R do stávajících postupů. Pomocí SPSS Custom Dialog Builder můžete R kód sdílet a předat uživatelům, kteří sami programovat nechtějí. Programátoři se tak mohou soustředit na práci s kódem a analytici na interpretaci analýz.

Komunita predikčních analýz AnalyticsZone

Díky oblibě nástroje SPSS Modeler je i komunita jeho uživatelů poměrně velká. Stanete-li se součástí komunity AnalyticsZone, získáte přístup ke zdroji, kde uživatelé mohou sdílet svou práci, vzájemně se ovlivňovat, učit se a stahovat tutoriály, příklady a další materiály. Rostoucí katalog rozšíření Predictive Extensions může být použit k rozšíření funkcionality prostředí SPSS Modeler. Jako člen komunity můžete jednoduše stahovat aktuální rozšiřující uzly nebo sem přidávat svoje vlastní.

Specifikace

Příprava dat

- přístup k datům z Cognos Business Intelligence, Cognos TM1, IBM DB2®, Oracle®, Microsoft SQLServer™, IBM Informix®, IBM Netezza, mySQL (Oracle) a datovým zdrojům Teradata
- import textových souborů pevné délky nebo s oddělovači, import datových souborů IBM SPSS Statistics, SAS, IBM SPSS Data Collection nebo XML
- přímý přístup k funkcionalitě správy dat a datovým transformacím z SPSS Statistics
- výběr proměnných, přejmenování, odvozování nových proměnných, kategorizace, nahrazení hodnot a změna pořadí proměnných
- výběr případů, náhodné výběry (včetně klastrového a stratifikovaného výběru), spojování dat (vnitřní, vnější, částečné vnější, inverzní) a textových řetězců, třídění, agregace a vážení
- datové restrukturalizace, rozdělení na testovací a tréninkovou množinu, transpozice
- funkce pro práci s textovými řetězci: tvorba řetězců, nahrazování znaků, vyhledávání, ořezávání a odebrání mezer
- RFM skórování (Recency, Frequency, Monetary), agregace transakčních dat pro kompletní RFM analýzu
- export dat do databází, IBM Cognos Business Intelligence, SPSS Statistics, SPSS Data Collection, textových souborů, Excel, SAS, XML
- Collection, textových dokumentů, Excel, SAS, XML.

Modelovací algoritmy

- Anomaly Detection: detekce neobvyklých záznamů
- Apriori: Asociační algoritmus s pokročilými funkcemi pro výběr pravidel
- Apriori Association Rules: Hledání asociací nakonfigurované pro big data

- *Bayesian Networks: Bayesovské sítě založené na podmíněné pravděpodobnosti*
- *C&RT, C5.0, CHAID & QUEST: Rozhodovací a klasifikační stromy s možností interaktivního růstu*
- *CARMA: Asociační algoritmus s možností vícenásobných důsledků*
- *Coxova regrese: Odhad realizace události ve specifikovaném čase*
- *Decision List: Interaktivní algoritmus pro vytváření rozhodovacích pravidel*
- *Factor/PCA: Algoritmy pro redukci dimenzionality*
- *K-Means a Two Step: seskupovací a segmentační algoritmy. KNN: Klasifikace metodou nejbližších sousedů*
- *Logistická regrese: Modelování binárních a nominálních proměnných*
- *Neuronové sítě: Vícevrstvá síť se zpětnou propagací, síť s radiální bazickou funkcí*
- *Regrese, lineární regrese, zobecněná lineární regrese (GLM) a zobecněné lineární smíšené modely (GLMM). Odhady parametrů v lineárních modelech*
- *Self-learning response model (SLRM): Naivní Bayesovské modely s postupným učením*
- *Sequence: Nalezení asociací v záznamech uspořádaných podle času*
- *SVM (podpůrné vektory): Pokročilý algoritmus vhodný pro rozsáhlé datové soubory*
- *TCM (Temporal Causal Modeling): Identifikace příčiny a důsledku vztahů v časových řadách nakonfigurované pro big data*
- *Time-Series: Analýza časových řad, automatické generování a odhady parametrů časových řad*

Modelování a evaluace

- *automatická klasifikace, regrese a seskupování pro rychlé nalezení vhodných modelů*
- *interaktivní prohlížeč modelů a přehledné statistické výstupy*
- *grafické zobrazení relativní důležitosti prediktorů pro predikci závislé proměnné*
- *kombinace několika modelů (metamodelování) nebo analýza jednoho modelu pomocí druhého*
- *Component-Level Extension Framework (CLEF) pro tvorbu vlastních uzlů*
- *přístup k nástrojům jazyka R pro rozšíření analytických možností*

Shrnutí

SPSS Modeler Personal umožňuje vytvářet predikční modely, najít vzory a trendy, skryté ve vašich strukturovaných datech a s vysokou mírou spolehlivosti predikovat budoucí vývoj. Ať už jste profesionální analytik nebo uživatel bez analytických znalostí, pokročilé predikční techniky vám pomohou dosáhnout těch nejlepších obchodních výsledků.