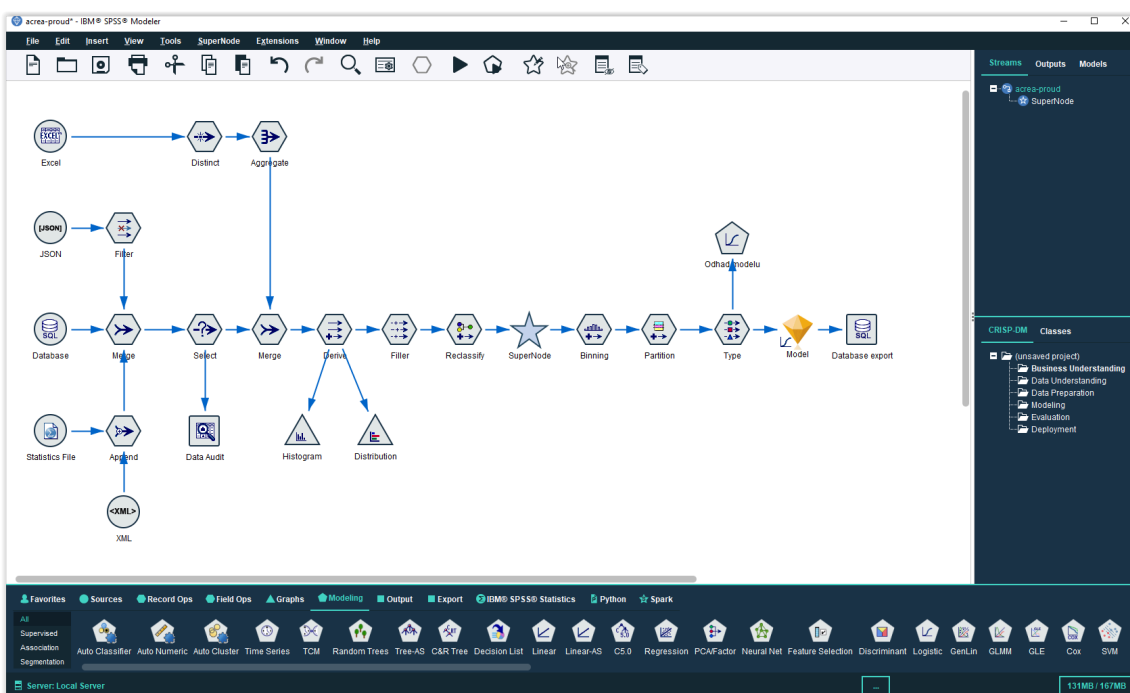


# PS CLEMENTINE

Komplexní, sofistikované, flexibilní, uživatelsky jednoduché, intuitivní a škálovatelné dataminingové řešení podporující celý analytický proces od přípravy dat přes modelování až po aplikaci modelů v praxi. Podporuje celý aplikační proces od správy analytických aktiv až po definici a historii jejich spuštění. Dokáže se přizpůsobit potřebám každé organizace. Vzhledem k univerzálnosti řešení najde uplatnění v různých odvětvích a oblastech podnikání a umožní řešit široké spektrum obchodních úloh i bez znalosti složitých analytických technik a technických aspektů jejich použití. Velice jednoduše jej lze integrovat do stávající IT infrastruktury a použít jako základ dalších řešení, například pro offline i online real-time skóring.

Řešení se skládá ze dvou částí. Dataminingová část je založena na jednom z nejrozšířenějších komerčních dataminingových nástrojů, IBM SPSS Modeler. Řešení obsahuje kompletní licenci s veškerou funkcionalitou, což zaručuje přístup k širokému spektru nástrojů, modelovacím algoritmům, statistickým a dataminingovým analýzám. IBM SPSS Modeler je systém založený na dataminingové metodologii CRISP-DM.



Druhá část má dva hlavní aspekty. Prvním aspektem je rozšířená funkcionalita v rámci dataminingového nástroje Modeler. Jedná se o přidané vstupní a procesní uzly, které mohou práci výrazně usnadnit. Jde například o uzly načítající sadu textových souborů se stejnou strukturou nebo uzel pro jednoduché i komplexní přejmenování proměnných. Dalším funkčním rozšířením je jednoduchá implementace volání webových služeb pomocí REST/SOAP rozhraní a to v jakékoli části datového proudu, tzv. streamu. Druhým aspektem je aplikační rozšíření. To představuje rozhraní pro sdílení a automatizaci spuštění dataminingových streamů. Toto rozšíření může mít dva módy. STANDALONE, kdy se vše instaluje na jeden počítač nebo režim WORKSTATION, kdy se jednotlivé komponenty aplikačního rozšíření mohou instalovat na různé počítače a které umožňuje v případě potřeby využít i serverovou část dataminingové aplikace IBM SPSS Modeler Server.

## PS CLEMENTINE PRO je ideální řešení když:

- potřebujete maximální výkon, ale serverové řešení je pro vás příliš robustní/nákladné
- spouštíte analytické úlohy ad hoc různými uživateli
- přistupujete k více datovým zdrojům (např. skladová data v relačních databázích a lokální souborová data v různých formátech) a potřebujete provést jejich načtení a konsolidaci
- automatizujete zpracování, generování a reportování dat (např. půlnoční spouštění při minimálním zatížení systémových prostředků a vytížení operačního systému)
- sdílíte analytické prostředky mezi mnoha procesy a uživateli

## PS CLEMENTINE PRO ve zkratce nabízí:

- statistické a dataminingové analýzy a prostředky pro přípravu dat v rozhraní klasického dataminingového nástroje
- integrované prostředí pro správu, uložení a přístup k analytickým prostředkům
- řízení procesu zpracování analytických úloh
- možnost týmové práce s uživateli s různými oprávněními
- ad hoc spuštění uložených, předdefinovaných a parametrizovaných analýz
- automatizované zpracování úloh podle časového plánu nebo na základě nějaké vnější systémové události
- možnost emailové notifikace o výsledku zpracování automatizovaných úloh či změně uloženého objektu
- monitoring zpracování analytických úloh

## Přidaná hodnota PS CLEMENTINE PRO:

- integrované prostředí pro správu analytických úloh
- prostředí pro uložení analytických zdrojů
- snadný přístup k analytickému obsahu
- kontrola a snadná správa analytických procesů
- zvýšená efektivita a komfort při provádění analytických úkolů
- úspora času při spouštění a zpracování úloh
- synergie během implementace analytických procesů

## PS CLEMENTINE PRO obsahuje tyto komponenty:

### IBM SPSS Modeler

Dataminingový nástroj používaný k exploraci dat a ke zpracování velkých objemů dat, poskytuje integraci s databázemi, poskytuje komplexní sadu technik pro strojové učení, supervizované i nesupervizované statistické metody a různé formy vizualizace výsledků.

### PS Desktop

Aplikace sloužící jako výchozí bod pro přístup k různé funkcionalitě PS CLEMENTINE, případně dalších produktů PS, například pro různou funkcionalitu nástroje PS Imago PRO.

### Webové rozhraní PS CLEMENTINE PRO

Rozhraní pro správu řešení a práci s uloženými analytickými aktivy - IBM SPSS Modeler streamy, případně jinými funkčními prvky, skripty. Zde se vytvářejí a nastavují automatizované úlohy.

”

*Výhody, které plynou z možnosti provádět komplexní operace s obrovským objemem dat jedinou osobou, která navíc přímo využívá výstupy a může je rychle a jednoduše implementovat, jsou neocenitelné.*

”

The screenshot shows the PS CLEMENTINE PRO web interface. The left sidebar contains navigation options: Last used, Active jobs, Event Log, Execution history, Search, Repository (with sub-items ACREA, Stream Execution, ONLY4MNGMT, and Stream Storage), Administration, My account, and Logout. The main area displays a table of folder contents for the ACREA repository. The table has columns for Name, Created, Modified, Versions, Type, In job, Next execution, Last execution, Last execution status, and Execution time. The data rows include various jobs and streams with their respective dates and execution statuses.

Name	Created	Modified	Versions	Type	In job	Next execution	Last execution	Last execution status	Execution time
psrest	19-07-2024 11:32	20-07-2024 09:21	2	Stream	✓				
samplebankloan	19-07-2024 11:32	19-07-2024 11:32	1	Stream	✓				
JOB2STREAMS	19-07-2024 11:34	06-08-2024 12:13	3	Job			06-08-2024 12:14	Failed	00:00:03
SOLAXREST	19-07-2024 11:36	19-07-2024 11:36	1	Service					
SOLAXDT84mins	19-07-2024 12:10	01-09-2024 07:53	3	Job		10-10-2024 11:00	10-10-2024 10:56	Executed	00:00:11
solax-PYE	05-08-2024 13:06	05-08-2024 13:06	1	Stream	✓				
spot	05-08-2024 13:06	05-08-2024 13:06	1	Stream	✓				
New job	05-08-2024 13:08	05-08-2024 13:12	1	Job					
Repository_Server_...	06-08-2024 12:15	06-08-2024 12:15	1						
Dokument	06-08-2024 12:16	06-08-2024 12:16	1						
UKAZKA	06-08-2024 12:59	06-08-2024 13:27	2	Stream	✓				
RTS-UKAZKA-M	06-08-2024 13:00	06-08-2024 13:17	1	Service					
RTS-UKAZKA-O	06-08-2024 13:06	06-08-2024 13:19	1	Service					

## PS CLEMENTINE PRO

Webové rozhraní PS CLEMENTINE PRO je jednotný nástroj pro správu řešení. Provádí se zde správa řešení ve smyslu správa uživatelských účtů, přiřazení rolí a práv k účtům a skupinám uživatelů. Spravuje se zde obsah, virtuální adresářová struktura a nastavují se práva k obsahu. Definují se zde joby, automatizované úlohy. V rámci jobů mohou být použity streamy IBM SPSS Modeler, které jsou v úložišti PS CLEMENTINE PRO uloženy, spravovány a verzovány. Kromě streamů se v automatizovaných úlohách dají použít i Python/PowerShell skripty.

Díky tomu, že je rozhraní webové, oproti starším verzím, nemusí se nikde dále již nic instalovat. Všem, administrátorům, power uživatelům ukládajícím streamy a definujícím joby, i business uživatelům, kteří si například chtějí pouze ad hoc spustit job, který jim vygeneruje nějaký report, tak stačí pouze webový prohlížeč. Různá úroveň přístupu a oprávnění k řešení PS CLEMENTINE PRO je řízena přidělením různým práv a rolí vybraným uživatelům či skupinám uživatelů.

PS CLEMENTINE PRO disponuje propracovaným systémem automatizace úloh. Úloha se může skládat z více funkčních prvků, typicky IBM SPSS Modeler streamů, které se spouští sekvenčně, nebo v rámci komplexnějšího procesu, který je definován "vývojovým diagramem". V rámci toho procesu se mohou spouštět i jiné funkční prvky, hlavně batch soubory z příkazové řádky serveru, případně jiné Python či PowerShell skripty. Díky tomu je možno automatizovat úlohy dalece přesahující dataminingový rámec IBM SPSS Modeler. Typicky se může jednat o úlohy, které nějakým způsobem operují s adresářovou strukturou nebo zpracovávají operace podmíněně stavem operačního systému, jiným aspektem vnějšího prostředí.

Obsah automatizovaných úloh, tzv. jobů je jeden z klíčových aspektů automatizace. Když je CO automatizovat, nabízí PS CLEMENTINE PRO celou řadu nástrojů na to KDY automatizovat.

The screenshot displays the PS CLEMENTINE PRO web interface. At the top, there's a navigation bar with the application name and user information. Below it, a table lists folder contents for 'OTHER', including jobs like 'pcrest', 'SOLAR10MiniREST', 'python', and 'PSC-UKAZKA'. The 'PSC-UKAZKA' job is highlighted, and its details are shown in a pop-up window. The 'Diagram' tab of this window shows a workflow starting from a 'Start' node, branching into 'pcrest' and 'python' tasks, which both lead to a 'CMB' task.

Name	Created	Modified	Versi...	Type	In job	Next execution	Last execution	Last execution status
pcrest	09-10-2024 15:01	09-10-2024 15:01	1	Stream	✓			
SOLAR10MiniREST	09-10-2024 15:02	10-10-2024 10:17	2	Job		11-10-2024 09:10	11-10-2024 09:06	Executed
python	10-10-2024 11:35	10-10-2024 11:35	1	Script	✓			
PSC-UKAZKA	10-10-2024 11:40	10-10-2024 11:43	1	Job		15-10-2024 12:00	10-10-2024 12:00	Executed

## Možnosti spouštění automatizovaných úloh

Připravené automatizované úlohy, joby, lze spouštět několika způsoby. Zdánilivě nejjednodušším způsobem je časovač. Ten umožňuje jít daleko nad klasické časování, např. každé ráno v 5 hodin. Umožňuje vybrat konkrétní dny v týdnu, první/poslední dny v měsíci, intervalové spouštění, např. každých 5 minut apod. Druhým způsobem spuštění jobu je zaslání "spouštěcí" zprávy pomocí protokolu REST/SOAP. PS CLEMENTINE PRO "naslouchá" a dostane-li pokyn ve formě webové zprávy, třeba i s dynamicky se měnícími parametry, tak spustí požadovaný job. Posledním způsobem spuštění je reakce na změnu v adresářové struktuře, tzv. file watcher. V případě, že v určeném adresáři přibudou nová data, např. nový soubor csv, tak se může spustit job.

## Výhody PS CLEMENTINE PRO a vybraná funkčnost

Prostřednictvím nezávislých studií a na základě našich i zákaznických zkušeností z dataminingových projektů jsme identifikovali potenciální výhody, které vám software může přinést:

- jednoduchá a rychlá implementace, nízké náklady na zaškolení v důsledku nenáročného ovládání softwaru,
- rychlé předání znalostí a know-how užívání softwaru v rámci organizace, zvýšení produktivity bez nutnosti programování
- nízké náklady na provoz, a to v krátkodobém i dlouhodobém horizontu
- nezávislost na datových zdrojích i objemu dat, ideální nejen pro tvorbu modelů, ale i pro přípravu dat
- nástroje pro automatizaci a sdílení predikčních modelů
- nízké TCO (Total Costs of Ownership – celkové náklady na vlastnictví) ve srovnání s klasickými nástroji IBM SPSS
- jednoduchá vizualizace datového procesu s širokou paletou nástrojů pro oblast datových manipulací a transformací
- datová otevřenost, uživatelská přívětivost a bohatá nabídka funkcí pro práci s daty
- možnost verzování proudů, jobů, uložených objektů v rámci Repository
- možnost automatizace spouštění proudů
- sdílení proudů ve společném úložišti v rámci analytického týmu