Erian GUHA Model Tester

Dokumentace aplikace pro zpětnou vazbu vytvořené v rámci TAČR TA04011691

# Získání zpětné vazby k chování data miningového modelu

Pro vyhodnocení chování data miningového modelu začleněného do business pravidla v systému Erian je v současné době využívána kombinovaná implementace

1. v interpretu data miningového modelu začleněného v komponentě *ExternalModel* vložitelné do těla business pravidla
2. v podobě samostatné aplikace *Erian GUHA Model Tester*

V rámci interpretu data miningového modelu v systému Erian je implementována podpora pro shromažďování informací o chování jednotlivých využitých asociačních pravidel. Dané informace jsou logovány v rámci úložiště systému Erian. Informacemi, které je možné následně analyzovat, jsou záznamy o:

* aktivacích jednotlivých pravidel (tj. splnění podmínky dané daným pravidlem)
  + „podmínkou pravidla“ je v případě využití modelu pro řešení klasifikačních úloh splnění podmínek definovaných antecedentem a podmínkou[[1]](#footnote-1) asociačního pravidla; v případě využití možnosti vracení měr zajímavosti dojde k aktivaci při splnění (pravdivosti) podmínek představovaných všemi částmi asociačního pravidla – tj. antecedentem, konsekventem i případnou podmínkou
* využití konkrétního pravidla jako výsledku vyhodnocení celého data miningového modelu
  + tj. dané pravidlo je preferovanou variantou z hlediska vyřešení konfliktních pravidel

## Data potřebná pro vyhodnocení zpětné vazby

Pro využití aplikace zpětné vazby je v současné verzi potřeba mít k dispozici data miningový model tvořený asociačními pravidly ve formátu GUHA PMML a export záznamu z úložiště (databáze) systému Erian ve formátu CSV.

Pro korektní vyhodnocení je nadále potřeba mít informaci o tom, jaký typ interpretu je využit v rámci ExternalModelu v business pravidle (tj. zda je cílem klasifikace/doporučování, či vracení hodnot měr zajímavosti).

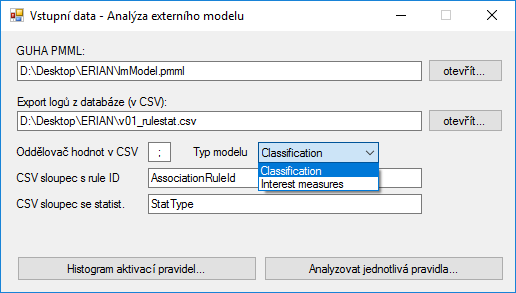
## Uživatelské rozhraní a ovládání aplikace

Aplikace disponuje jednoduchým uživatelským rozhraním, v rámci kterého uživatel nejprve vybere soubory s potřebnými daty a zadá potřebné parametry pro vyhodnocení – viz obrázek 1.

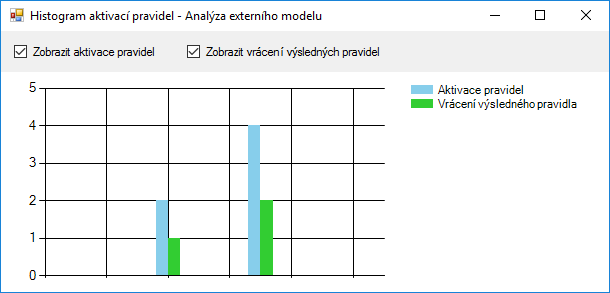
Následně má uživatel k dispozici dvě varianty zobrazení výsledků:

* zobrazení v podobě histogramu (viz obrázek 2)
* zobrazení v tabulkové podobě (viz obrázek 3)

Varianta zobrazení v podobě histogramu je výhodná pro rychlý přehled o aktivacích jednotlivých pravidel a pro možnost vizuálního porovnání rozložení četností aktivací/vrácení výsledných pravidel v rámci vyhodnocování dat z různých časových období. Při zastavení myši nad konkrétním „sloupcem“ v histogramu je uživateli zobrazena informace o konkrétním asociačním pravidle, kterého se daný sloupec týká.

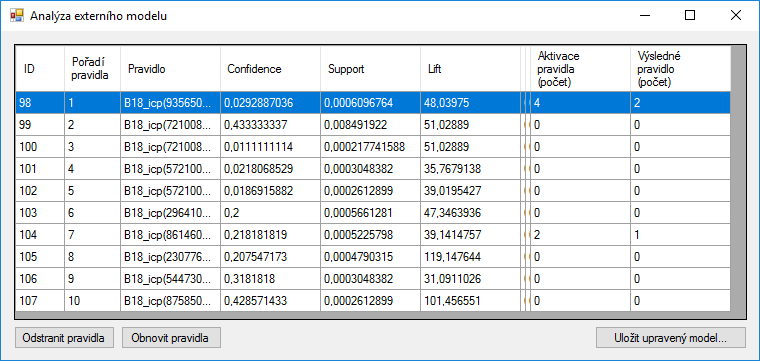


Obrázek 1 Aplikace Erian GUHA Model Tester - zadání vstupních dat



Obrázek 2 Aplikace Erian GUHA Model Tester – zobrazení logů v podobě histogramu

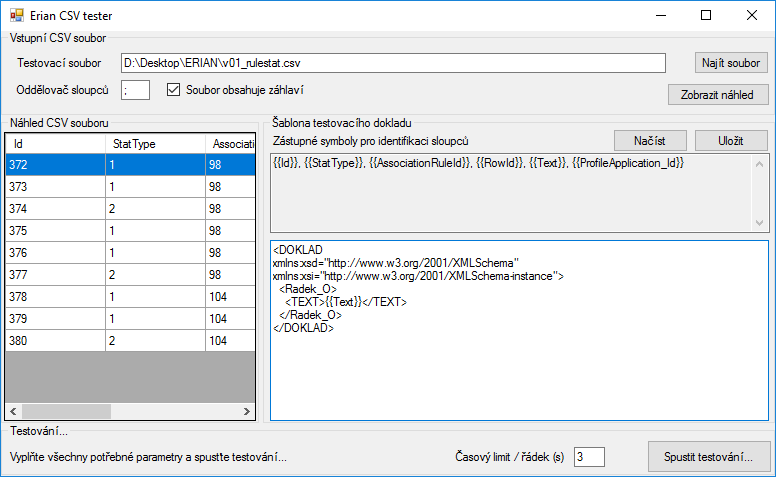
Zobrazení v podobě tabulky je výhodné pro podrobnější analýzu chování jednotlivých asociačních pravidel. V tomto zobrazení má uživatel možnost upravovat využívaný data miningový model prostřednictvím odstranění vybraných pravidel. Po provedené úpravě modelu je při kliknutí na tlačítko „Uložit upravený model…“ výsledný model uložen opět v podobě GUHA PMML. Lze ho tedy nahrát do komponenty ExternalModel v business pravidle v systému Erian, které se následně začne chovat dle upraveného modelu.



Obrázek 3 Aplikace Erian GUHA Model Tester – zobrazení logů v podobě tabulky

# Vyhodnocení dat využitých k data miningu systémem Erian

Pro jednoduché hromadné otestování dat využitých pro dolování má uživatel k dispozici jednoduché rozhraní (viz obrázek 4) pro možnost odeslání dat webové službě systému Erian. Pro otestování uživatel vybere vstupní data v podobě CSV soubor a definuje vzor odesílaných dat. Odesílaná data (v názvosloví systému Erian „doklad“) jsou ve vstupním CSV souboru představována jedním řádkem, v odesílané podobě poté XML strukturou v závislosti na typu daných dat. Pro definování vzoru XML struktury jsou využívány symbolické zástupné symboly představující hodnoty jednotlivých sloupců odvozené od jejich názvů či od jejich pořadí v CSV souboru. Např. *{{Id}}*.



Obrázek Rozhraní pro odeslání dat využitých pro data mining pro vyhodnocení systémem Erian

# Implementace

V současné verzi je aplikace *Erian GUHA Model Tester* implementována v podobě samostatné aplikace. Aplikace je vytvořena v jazyce C# pro rozhraní .NET Framework 4.5.2. V rámci implementace je kromě standartních prostředků daného rozhraní využita také komponenta CsvHelper.[[2]](#footnote-2)

Výsledkem kompilace projektu pro zpětnou vazbu je dvojice samostatných spustitelných souborů – hlavní část aplikace zaměřená na zobrazování zpětné vazby (ErianDbLogs.exe) a aplikace pro spuštění vyhodnocení dat využitých pro data mining přes standartní webovou službu systému Erian (ErianCsvTester.exe)

1. GUHA asociační pravidla se vždy skládají z antecedentu a konsekventu, volitelně však může být využita také třetí část, představující omezující podmínku [↑](#footnote-ref-1)
2. https://joshclose.github.io/CsvHelper/ [↑](#footnote-ref-2)