Narzędzie eksploracji danych

Kornel Lewandowski, Małgorzata Napieraj

Opis projektu

Proste w obsłudze narzędzie do eksploracji danych medycznych, które umożliwia tworzenie klasyfikatorów dla zawartych w danych wzorców, wstępne przetwarzanie danych i atrybutów, import i eksport danych do popularnych formatów plików oraz łatwy w obsłudze i przyjazny interfejs.

Koncepcja realizacji

Aplikacja działająca na wielu platformach

Przejrzysty interfejs oparty o przeglądarkę

Nie chcemy odkrywać koła na nowo

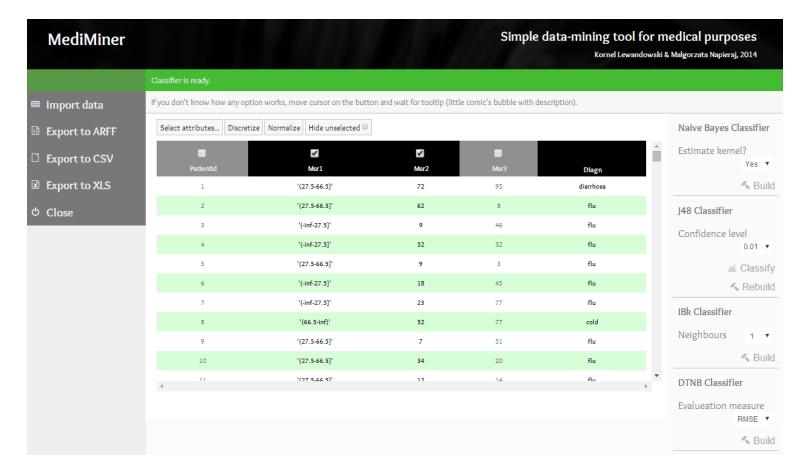
Specyfikacja

Lp	Opis funkcjonalności	Тур	Status
1	Normalizacja i dyskretyzacja danych	obowiązkowa	
2	Selekcja atrybutów	obowiązkowa	
3	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>NaiveBayes</i>)	obowiązkowa	
4	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm J48)	obowiązkowa	
5	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>IBk</i>)	obowiązkowa	
6	Klasyfikacja przykładów na zbudowanym klasyfikatorze	obowiązkowa	
7	Manipulacja parametrami algorytmów klasyfikacji	obowiązkowa	
8	Import danych w formacie XLS/CSV/ARFF	obowiązkowa	
9	Eksport danych w formacie XLS/CSV/ARFF	obowiązkowa	
10	Pomoc kontekstowa	obowiązkowa	
11	Import/eksport modelu klasyfikatora	opcjonalna	
12	Import danych w formacie XML	opcjonalna	
13	Eksport danych w formacie XML	opcjonalna	
14	Udostępnienie dodatkowego algorytmu selekcji atrybutów (GainRatio)	opcjonalna	
15	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm DTNB)	opcjonalna	

Stopień realizacji

Lp	Opis funkcjonalności	Тур	Status
1	Normalizacja i dyskretyzacja danych	obowiązkowa	100,00%
2	Selekcja atrybutów	obowiązkowa	100,00%
3	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>NaiveBayes</i>)	obowiązkowa	100,00%
4	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm J48)	obowiązkowa	100,00%
5	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>IBk</i>)	obowiązkowa	100,00%
6	Klasyfikacja przykładów na zbudowanym klasyfikatorze	obowiązkowa	100,00%
7	Manipulacja parametrami algorytmów klasyfikacji	obowiązkowa	100,00%
8	Import danych w formacie XLS/CSV/ARFF	obowiązkowa	100,00%
9	Eksport danych w formacie XLS/CSV/ARFF	obowiązkowa	100,00%
10	Pomoc kontekstowa	obowiązkowa	100,00%
11	Import/eksport modelu klasyfikatora	opcjonalna	0,00%
12	Import danych w formacie XML	opcjonalna	0,00%
13	Eksport danych w formacie XML	opcjonalna	0,00%
14	Udostępnienie dodatkowego algorytmu selekcji atrybutów (GainRatio)	opcjonalna	0,00%
15	Budowa klasyfikatora na podstawie zbioru uczącego (algorytm <i>DTNB</i>)	opcjonalna	100,00%

Interfejs



MediMiner a WEKA Explorer - różnice

Przyjazny interfejs

Pomoc kontekstowa

Podgląd danych

Import/eksport danych w formacie XLS

Napotkane problemy

UTF8 ← ANSI: polskie znaki

Normalizacja - wszystkie atrybuty

Atrybut decyzyjny typu liczbowego

Dyskretyzacja atrybutu decyzyjnego

Instancje: reprezentacja danych oraz ich etykiet

Co dalej?

Odczyt/zapis modelu klasyfikatora
Odczyt/zapis wyników klasyfikacji
Dodawanie/usuwanie instancji ze zbioru
treningowego

Udostępnienie dodatkowych funkcji WEKI

Technologie

```
Java
  Maven
  Weka
HTML + CSS
JavaScript
  jQuery
  Jetty
```

Prezentacja aplikacji