

Sztuczna inteligencja w finansach

Projekt 2

Data oddania projektu: 13 kwietnia 2016

Celem projektu jest opanowanie metod poszukiwania wiedzy baz danych finansowych z wykorzystaniem sieci neuronowych i eksploracji danych. Twoim zadaniem jest opracowanie modelu klasyfikującego klientów do odpowiednich grup ryzyka na podstawie danych historycznych i ich sytuacji finansowej. Taki model będzie zastosowany w ocenie wiarygodności finansowej klienta przy udzielaniu kredytu, wykupieniu ubezpieczenia oraz wyznaczaniu strategii windykacyjnej.

Opracowany model ma klasyfikować klientów do jednej z następujących grup ryzyka oceny sytuacji finansowo-kredytowej klienta: Low (L) - oznacza ryzyko niskie, Medium (M) – ryzyko średnie, High (H) – ryzyko wysokie, High Plus (HP) – ryzyko bardzo wysokie. Na podstawie klasyfikacji będzie wybrana najbardziej odpowiednia strategia windykacyjna (sposób postępowania) dla danego klienta. Model oceny sytuacji finansowej ma też służyć do interpretacji/ wyjaśnienia decyzji windykacyjnych na podstawie oceny ryzyka sytuacji finansowej dłużnika.

Wyznaczenie w/w grup klientów i podanie ich charakterystyk należy dokonać z wykorzystaniem metod eksploracji danych; rekomenduję platformę finansową Pathfinder firmy VSoft. W analizie można też użyć innych pakietów eksploracji danych takich jak: Orange, R, Weka, Statistica, SAS,.. W celu sklasyfikowania obserwacji można skorzystać z dostępnych algorytmów (np. w zakładce *Classify* w Orange), takich jak drzewa decyzyjne, MLP, Naive Bayes, CN2,..... W prezentacji projektu należy skomentować otrzymane reguły i przedstawić wizualizację procesu eksploracji oraz wyników. Należy też ocenić jakość klasyfikatora.

Z uwagi na ograniczenia związane z licencją, Platforma jak i dane będą dostarczone imiennie.

Wskazówki:

- A. Zapoznać się z platformą Pathfinder (patrz zał. Instrukcja)
- B. Wykorzystać Excel do transformacji danych źródłowych
- C. Zainstalować i zapoznać się z działaniem platformy Orange (<http://orange.biolab.si/>).
- D. Przygotować i przesłać prezentację Power Point jak i skompresowany plik o nazwie **SIwF-nazwisko.zip** na adres: jerzy.korczak@ue.wroc.pl
Prezentacja powinna zawierać: plik PowerPoint pokazujący metodę rozwiązania problemu, uzasadnienie wyboru modelu i zmiennych, dane eksperymentalne, schemat procesu eksploracji, wizualizacje i konkluzje z przeprowadzonych badań.
- E. Projekt może być zrealizowany przez max dwóch studentów.

Pomoce:

- Materiały z wykładów i laboratoriów (<http://www.korczak-leliwa.pl>)
- Instrukcja Pathfinder
- Orange Tutorial