

Elementy Sztucznej Inteligencji - sztuczne sieci neuronowe.

Mateusz Jakubczak, Krzysztof Olipra, Karol Oleszek

Temat projektu:

Czy klient założy lokatę? - klasyfikacja binarna w oparciu o dane telemarketingowe

Struktura projektu:

- bank.csv dane z UCI
- compare_methods.py porównanie z innymi metodami
- compare_methods_report.txt dokładny raport z porównania

Dane pochodzą z otwartego repozytorium zbiorów danych do uczenia maszynowego.

Link: <https://archive.ics.uci.edu/ml/datasets/Bank%2BMarketing>

Użyte dane pochodzą z działań marketingowych anonimowego Portugalskiego banku. Kampanie marketingowe opierały się na telefonach do klientów.

Celem klasyfikacji jest przewidzenie, czy klient założy lokatę po telefonie telemarketera.

Dane o kliencie:

- wiek
- praca
- stan cywilny
- edukacja
- czy jest bankrutem
- czy ma hipotekę
- czy ma pożyczki

Dane o ostatnim kontakcie z telemarketerem:

- czy telefon stacjonarny
- miesiąc kontaktu
- dzień tygodnia kontaktu
- czas rozmowy

Inne zmienne:

- liczba poprzednich telefonów do klienta
- dni od poprzedniego kontaktu
- sukces poprzednich kontaktów

Dane makroekonomiczne:

- wariancja kwartalnego bezrobocia
- CPI miesięczne
- Consumer Confidence Index
- stopa euribor 3
- liczba zatrudnionych w gospodarce

S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimaraes, Portugal, October, 2011. EUROSIS. [bank.zip]

W podanych źródłach autorzy analizują strukturę zbioru danych, metody doboru zmiennych oraz porównują efektywność różnych technik uczenia maszynowego. Autorzy najwyższą efektywność osiągnęli przy użyciu Support Vectors Machines. Opisane są również kroki niezbędne do efektywnego użycia modeli w środowisku biznesowym, m. in. poprzez wyjaśnialność modeli.

Analiza wpływu - liczba warstw

Analiza wpływu - liczba neuronów

Analiza wpływu - rodzaj funkcji aktywacji

Do porównania wybraliśmy modele z Scikit-learn:

- Drzewo decyzyjne
- Naive Bayes
- K-najbliższych sąsiadów
- Support Vector Machines

Porównanie wyników - inne metody

Metoda	Poprawność(Accuracy)
SSN	x%
Drzewo decyzyjne	68%
Naive Bayes	68%
KNN	60%
SVC	58%

Pełen raport z klasyfikacji w pliku `compare_methods_report.txt`