Projekt – Uczenie maszynowe

Piotr Kondratowicz

# Cel projektu

Celem projektu było stworzenie modelu, który przewiduje, czy klienci pewnego hotelu, który przeszedł drastyczne zmiany uszanują lub odwołają rezerwację.

# Dane

Dane pochodzą z platformy kaggle.com (link: [Hotel Reservations Dataset | Kaggle](https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset)) i zawierają informacje o różnych rezerwacjach z cechami tj. liczba dorosłych i dzieci, cena pokoju, liczba zarezerwowanych dni, data przyjazdu czy plan posiłków.

Dane składają się z 36275 przykładów, które podzielono na zbiór uczący (32637 przykładów) i zbiór testowy (3628 przykładów).

# Modele

W projekcie porównano działanie 3 modeli:

* Regresja logistyczna. Jako optymalizatora użyto Stochastic Gradient Descent (SGD). Zastosowano regularyzację L2.
* Algorytm k-najbliższych sąsiadów dla k = 13. Jako metrykę odległości zastosowano odległość Manhattan.
* Sieć neuronowa zbudowana z 5 warstw ukrytych z funkcją aktywacji „relu” oraz końcowa z funkcją aktywacji „sigmoid”. Zastosowano optymalizator „Adam” oraz funkcję kosztu „Binary Crossentropy”. Ilość parametrów: 157297, ilość epok: 100.

# Ewaluacja

Do ewaluacji wykorzystano metryki: accuracy, precision, recall i F1-score. Wyniki ewaluacji na zbiorze testowym przedstawia poniższa tabelka:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-score |
| Regresja logistyczna | 79% | 77% | 73% | 74% |
| Algorytm k-najbliższych sąsiadów | 86% | 85% | 83% | 84% |
| Sieć neuronowa | 88% | 87% | 87% | 87% |

# Wnioski

Najlepsze wyniki pod względem każdej metryki uzyskała sieć neuronowa. Drugie miejsce uzyskał algorytm k-najbliższych sąsiadów tracąc do sieci neuronowej średnio 3 punkty procentowe na każdej metryce. Najgorsze wyniki zostały uzyskane przy użyciu regresji logistycznej, która traci średnio 9 punktów procentowych względem algorytmu k-najbliższych sąsiadów i aż 12 względem sieci neuronowej. Nie ulega wątpliwości, że sieć neuronowa oraz algorytm k-najbliższych sąsiadów poradziły sobie całkiem dobrze z rozwiązaniem problemu rezerwacji hotelowych.