Wzór raportu z projektu

06-DUMAUI0 2021/SL

# Cel projektu

Celem projektu było stworzenie modelu, który przewiduje, czy lokal znajduje się w centrum miasta na podstawie jego powierzchni, ceny i typu zabudowy.

# Dane

Dane pochodzą z wyzwania „Mieszkania 5” na platformie Gonito.pl (link: https://gonito.net/challenge/mieszkania5).

Po odrzuceniu obserwacji odstających, uzyskano 4980 przykładów, które podzielono na zbiór uczący (3984 przykłady) i zbiór testowy (996 przykładów).

# Modele

W projekcie porównano działanie 3 modeli:

* Regresja logistyczna wielomianowa 3. stopnia. Jako optymalizatora użyto mini-batch gradient descent z wielkością batcha równą 100. Zastosowano regularyzację L2.
* Regresja logistyczna wielomianowa 3. stopnia. Jako optymalizatora użyto mini-batch gradient descent z wielkością batcha równą 100. Brak regularyzacji.
* Naiwny klasyfikator bayesowski. Jako modelu klas użyto rozkładu normalnego.

# Ewaluacja

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-score |
| Regresja logistyczna z regularyzacją |  |  |  |  |
| Regresja logistyczna bez regularyzacji |  |  |  |  |
| Naiwny klasyfikator bayesowski |  |  |  |  |

# Wnioski

Najlepsze wyniki pod względem F1-score uzyskano przy pomocy naiwnego klasyfikatora bayesowskiego. Model regresji logistycznej bez regularyzacji poradził sobie nieco gorzej niż model regresji logistycznej z regularyzacją, co prawdopodobnie wynika ze zjawiska nadmiernego dopasowania.