Wstep do uczenia maszynowego
Raport walidacyjny

Michał Iwaniuk, Bartosz Jezierski

1 Wstep

Naszym zadaniem jako walidatorów projektu *New York Housing Market*, prowadzonego przez Krzysztofa Adamczyka i Pawła Florka, było ostateczne przetestowanie przez nich zaproponowanych modeli oraz monitorowanie ich pracy przez cały okres projektu. Głównym celem zadania było sklasyfikowanie cen mieszkań w Nowym Jorku.

2 Proces walidacyjny

Po ukończeniu każdego z etapów przez grupe modelarzy otrzymywaliśmy do wgladu ich postep w pracy i zwracaliśmy im nasza ocene oraz ewentualne sugestie, co należałoby poprawić.

2.1 Dane walidacyjne

Na samym poczatku prac otrzymaliśmy zestaw danych walidacyjnych, który składał sie z 30% oryginalnego zbioru danych. Grupa modelarzy nie miała dostepu do tych danych podczas tworzenia modelu, a naszym zadaniem było wykorzystanie ich do ostatecznej oceny modelu. Poczatkowo modelarze zapomnieli podzielić zbioru na zbiór modelarzy i walidatorów oraz nie ustalili ziarna losowości, jednak szybko to poprawili.

2.2 EDA

Grupa modelarzy poradziła sobie bez wiekszych zastrzeżeń. Przebadali oni zmienne kategoryczne oraz numeryczne w odpowiedni sposób. Wykonali wszystkie podstawowe kroki, takie jak: badanie rozkładów zmiennych numerycznych, analize relacji miedzy zmiennymi, ich korelacji oraz zależności miedzy tymi zmiennymi a wartościa docelowa (cena). Przebadali również zmienne kategoryczne, zliczajac ich wartości, unikalne wartości oraz badajac ich zależności od ceny. Modelarze wyciagneli poprawne wnioski.

2.3 Feature engineering

W tym etapie również wykonano wszystkie potrzebne kroki. Modelarze w logiczny sposób usuneli z ramki niepotrzebne kolumny, które zazwyczaj nie niosły żadnej wartości nie zawartej w pozostałych kolumnach. Mieliśmy jednak zastrzeżenia co do usuniecia kolumn zawierajacych informacje o szerokości i długości geograficznej danego mieszkania, które de facto określaja jego położenie, co może mieć wpływ na cene. Modelarze zastosowali sie do naszych sugestii, co przyczyniło sie do poprawy wyników. Usuneli również outliery z ramki i przekształcili odpowiednio zmienne numeryczne. Dla zmiennych kategorycznych wszystkie pojedyncze obserwacje wrzucili do kategorii *Other*, aby zmniejszyć ich ilość. Nastepnie zaenkodowali zmienne kategoryczne, poczatkowo przy użyciu OneHotEncoder, jednak po naszych sugestiach zamienili go na TargetEncoder, ponieważ zmiennych kategorycznych było za dużo.

2.4 Hiperparametry

W tym etapie modelarze przedstawili wyniki uzyskane dla różnych modeli, w tym bardziej zaawansowanych modeli takich jak XGBClassifier. Do ustalenia hiperparametrów wykorzystali metode GridSearchCV oraz RandomizedSearchCV oraz przeprowadzili walidacje krzyżowa. Nie mieliśmy wiekszych zastrzeżeń na tym etapie. Wyniki porównali również z metoda AutoML, które były podobne, co pokazuje, że modelarze poradzili sobie poprawnie z zadaniem. Aczkolwiek użyli jedynie 3 generacje, przez co tpot mógł niewystarczjaco sie zoptymalizować, na co zwróciliśmy im uwage.

3 Podsumowanie

Podsumowujac, rozwiazanie zaproponowane przez grupe modelarzy działa poprawnie na danych, które nie zostały wykorzystane do jego stworzenia. Uważamy, że z dostepnych danych wyciagneli bardzo zbliżone wnioski do naszych oraz uzyskali wyniki zbliżone do wyników na naszym zbiorze danych.

3.1 Porównanie wyników

Oto przykładowe wyniki uzyskane przez modelarzy oraz walidatorów dla modelu RFC, jak widać sa bardzo zbliżone.

3.1.1 modelarze

```
RFC:
mean: 0.8544768288716361, std: 0.012921780071181489
0.8552631578947368
              precision    recall  f1-score   support

           0       0.93      0.76      0.84       146
           1       0.83      0.95      0.89       360
           2       0.85      0.67      0.75       102

    accuracy                           0.86       608
   macro avg       0.87      0.79      0.82       608
weighted avg       0.86      0.86      0.85       608
```

3.1.2 walidatorzy

```
RFC:
mean: 0.8525844915318601, std: 0.018025627239877384

0.8435114503816794

              precision    recall  f1-score   support

           0       0.81      0.77      0.79        61
           1       0.87      0.90      0.88       171
           2       0.77      0.67      0.71        30

    accuracy                           0.84       262
   macro avg       0.81      0.78      0.80       262
weighted avg       0.84      0.84      0.84       262
```