

Projekt - Deferred Learning

Olaf Werner, Karol Saputa

12 czerwca 2019

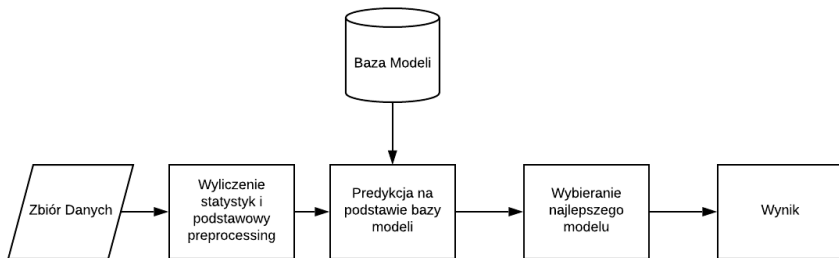
Zbiór danych

W naszym zbiorze danych mieliśmy mieliśmy 23 kolumny, ale duża część z nich się powtarzała więc w ramach preprocessingu usuwaliśmy je. Po za tym kolumny w które miały tylko dwie unikalne wartości traktowaliśmy jako zmienne logiczne. Kolumnę Emilia potraktowaliśmy logarytmem, a kolumnę Anna potraktowaliśmy jako factor, a potem one hot encode.

Zastosowana metoda

By oszacować najlepszy możliwy model dla naszego zbioru danych wyliczyliśmy dla niego statystyki takie jak ilość zmiennych numerycznych, wielkość zbioru itd. po czym używaliśmy modelu predycyjnego zbudowanego na naszej bazie modeli by znaleźć dla niego najlepszy model.

Zastosowana metoda



Szukanie ulepszeń

- ▶ Budowa bardziej zaawansowanego metamodelu
- ▶ Zmiany w części dotyczącej preprocessingu
- ▶ Dobór hiperparametrów

Wykorzystanie sprawdzania etykiet

- ▶ Powtarzające się rekordy w zbiorze testowym po zweryfikowaniu okazały się mieć różne klasy.
- ▶ Zastanowiliśmy się czy możemy zrobić coś innego niż zwykle usunięcie tych obserwacji.
- ▶ Zdecydowaliśmy się pozostawić po jednej obserwacji w zbiorze przy zastosowaniu mody dla etykiet.

Wykorzystanie hiperparametrów

- ▶ Na podstawie omawianej pracy ‘Tunability: Importance of Hyperparameters of Machine Learning Algorithms’ można znaleźć ogólnie dobre hiperparametry.
- ▶ Można wykorzystać tą metodę dla modeli wybieranych przez metamodel (innych niż te opisane w pracy). Na podstawie predykcji na zbiorze testowym.
- ▶ Niestety nie udało się osiągnąć oczekiwanych rezultatów ze względu na błędy oraz długość obliczeń.

Wnioski

- ▶ Najważniejszy jest preprocessing. Ma on wpływ na wynik metamodelu.
- ▶ To jest przyszłość machine learning, ponieważ pozwala automatyzować pracę.
- ▶ Jednak w naszym przypadku na przykład brak uwzględnienia w metamodelu hiperparametrów prowadził do wyboru przez metamodel takich modeli, które po prostu ogólnie dobrze sobie radziły. W praktyce, na innych zbiorach, łatwo można było znaleźć modele lepsze.
- ▶ Metamodel doboru hiperparametrów?