

Projekt - Deferred Learning

Olaf Werner, Karol Saputa

23 kwietnia 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 predykcyjnego zbudowanego na naszej bazie modeli by znaleźć dla niego 5 najlepszych modeli, porównywaliśmy też z najbardziej skutecznymi i popularnymi modelami i wybieraliśmy najlepszy z nich .

Zastosowana metoda

