

Отчет

df_buffer_table1 – таблица с таргетом и маской разбиения на трейн-тест (идентификатор - cash_hist)

df_buffer_table2, df_buffer_table3 – таблицы со статистикой (идентификатор - REQUEST_ID)

1. таблицы df_buffer_table2, df_buffer_table3 были агрегированы в разрезе REQUEST:
2. мешки категорий
3. среднее, минимум, максимум, доля нулей
4. разбили на бины промежутки между таймстемпами (1 мин, 2 мин и тд) и собрали из них мешки
5. применили tfidf к paymtPat

жадным образом отобрал признаки

измерил валидационную точность и точность на отложенной выборке: они сильно разнились из-за различий в распределениях

с помощью бутстрепа оценил переобучение и сделал блендинг предсказаний для тестовой выборки