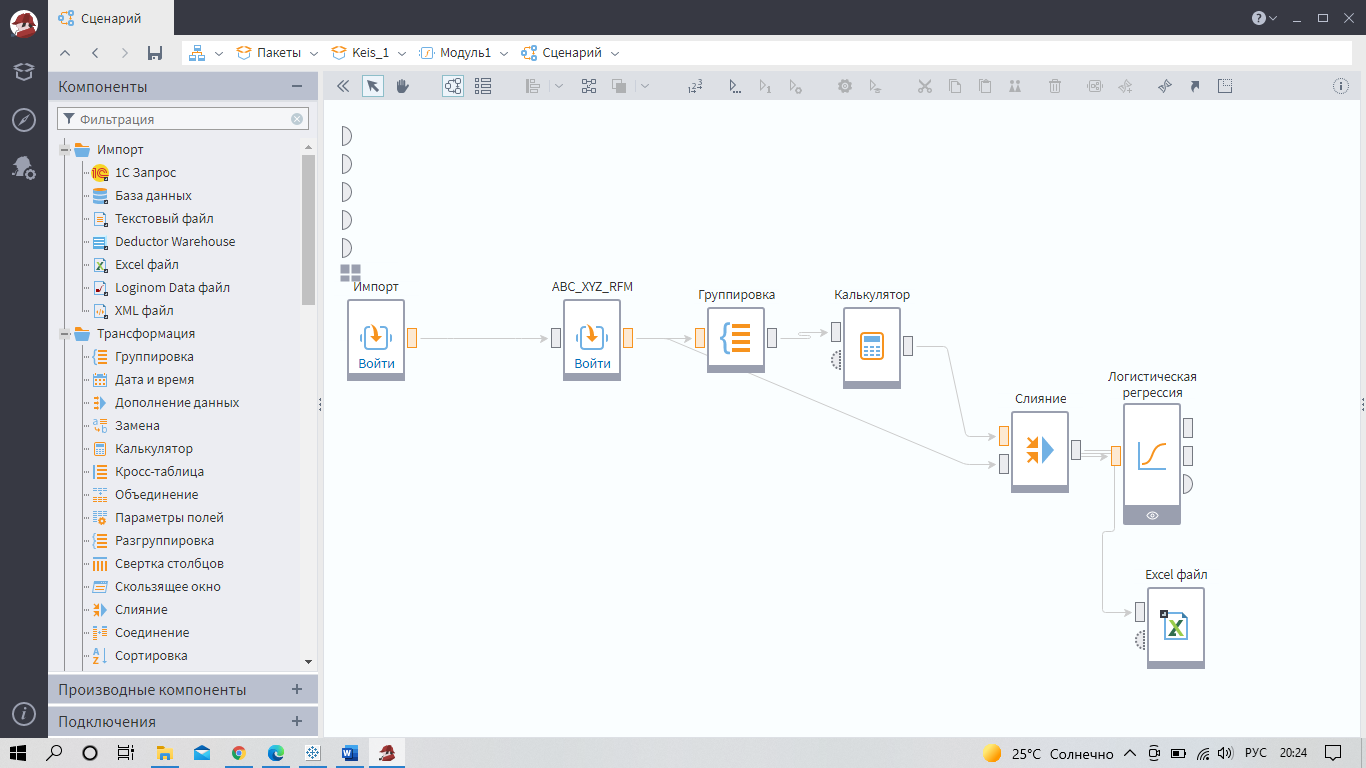
В качестве данных для аттестационной работы был взят кейс №1.

Было дано четыре таблицы с данными о продукте, клиентах, территории продаж и самих продажах.

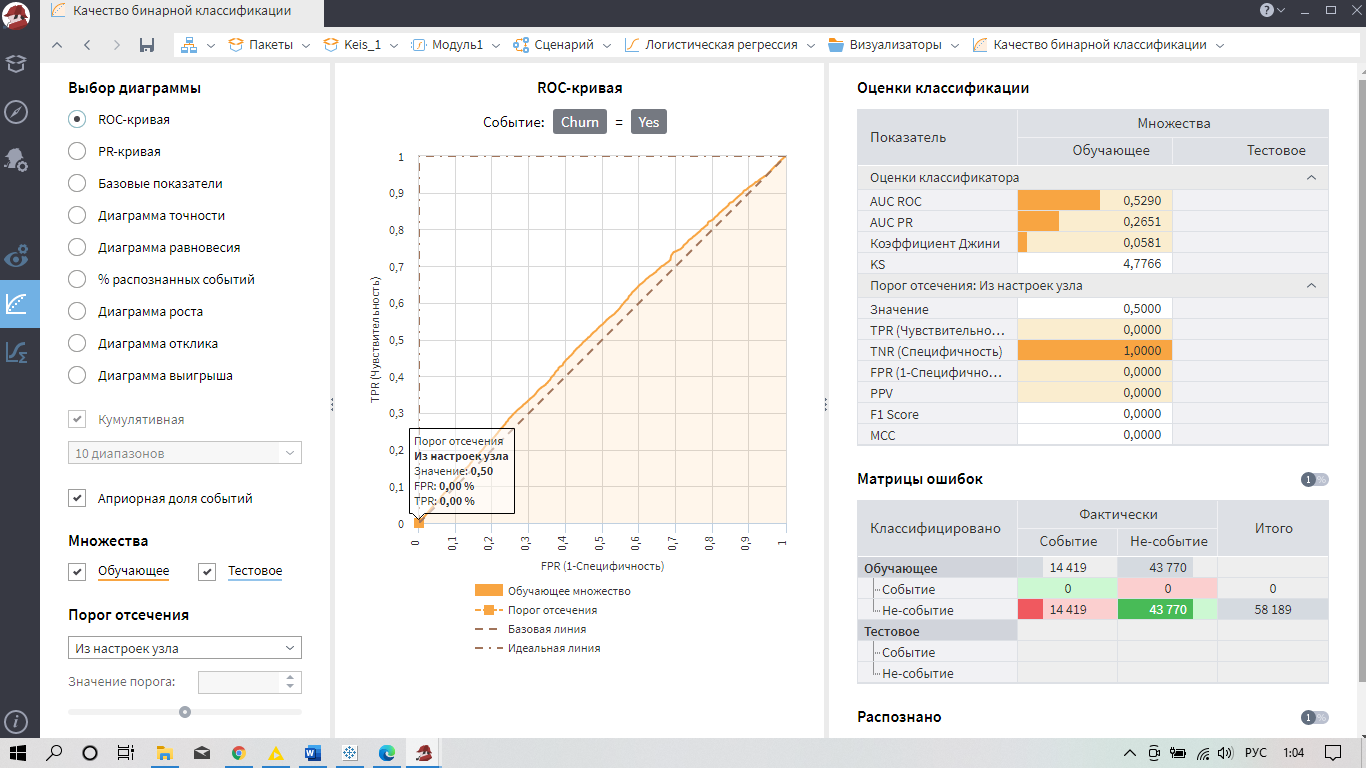
Продуктами являлись байки, специализированная одежда и аксессуары. Продажи осуществлялись на территории США, Австралии и Европы.

Объединение таблиц и очистка данных производилось в платформе Loginom.

Там был выполнен ABC-XYZ и RFM анализ. Так же был создан новый столбец – отток клиентов (Churn). В Loginom была решена задача классификации с помощью логистичесой регрессии.

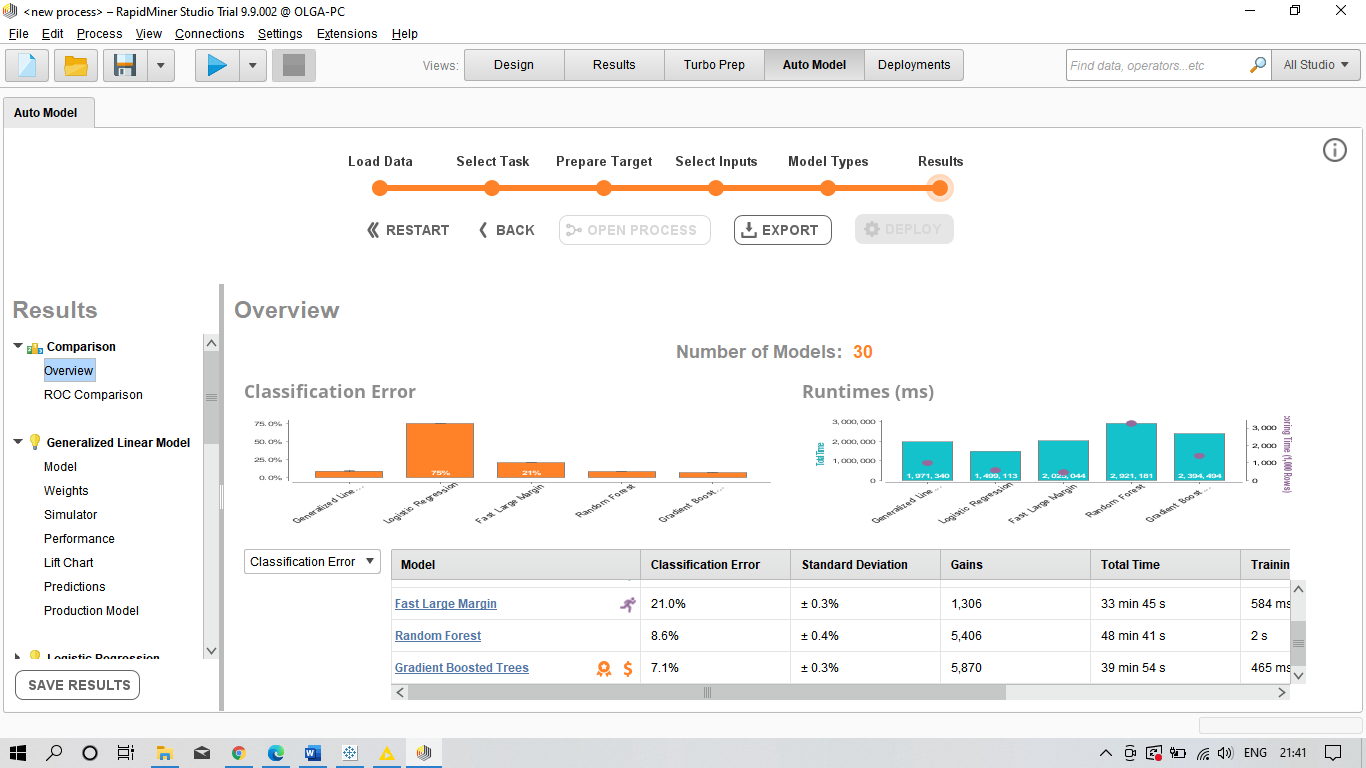


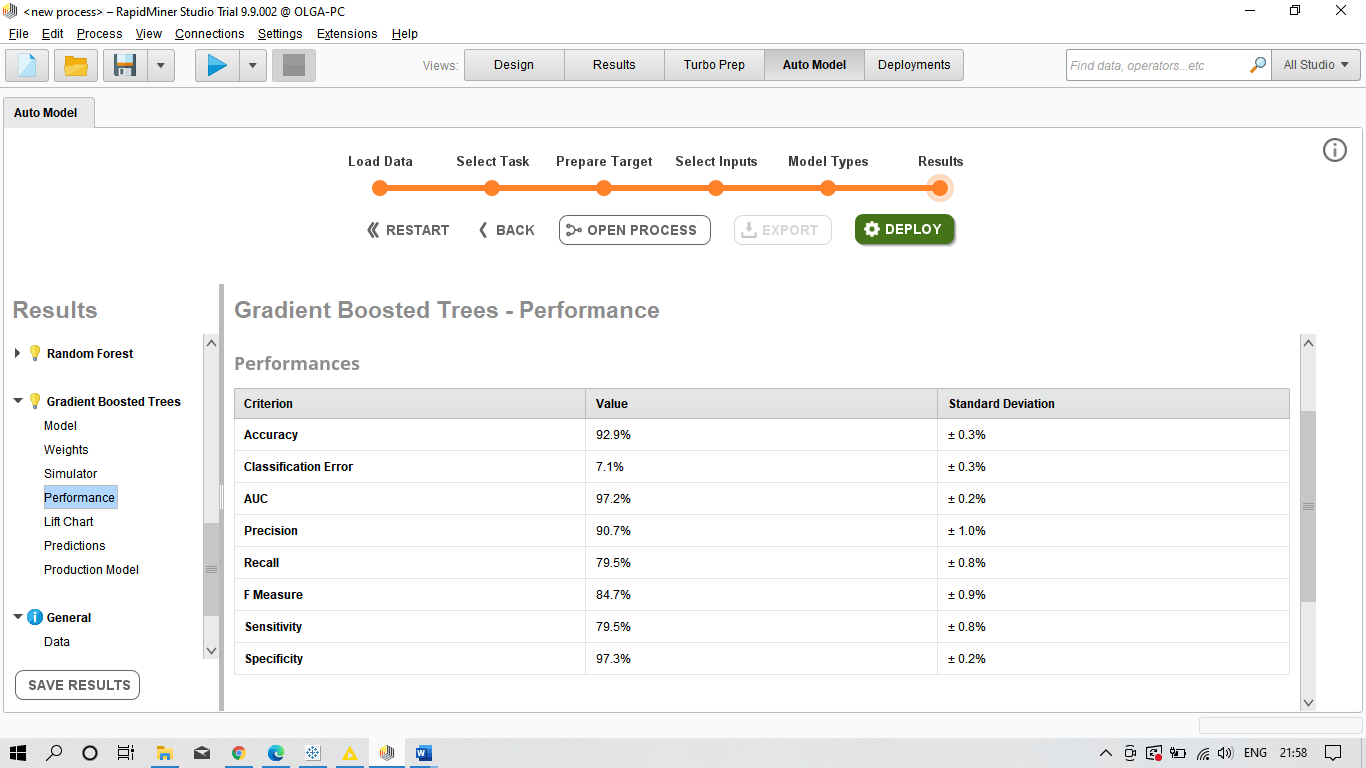
Логистическая регрессия, реализованная при помощи кроссвалидации в Loginom:



Логистическая регрессия дала плохой результат: ошибки 50%.

Поэтому задача классификация была решена в платформе RapidMiner:





Программа дала отличные результаты: Ошибки 7%, AUC 97%. Лучшей моделью стал Градиентный бустинг с такими параметрами: количество деревьев 10, глубина 6.

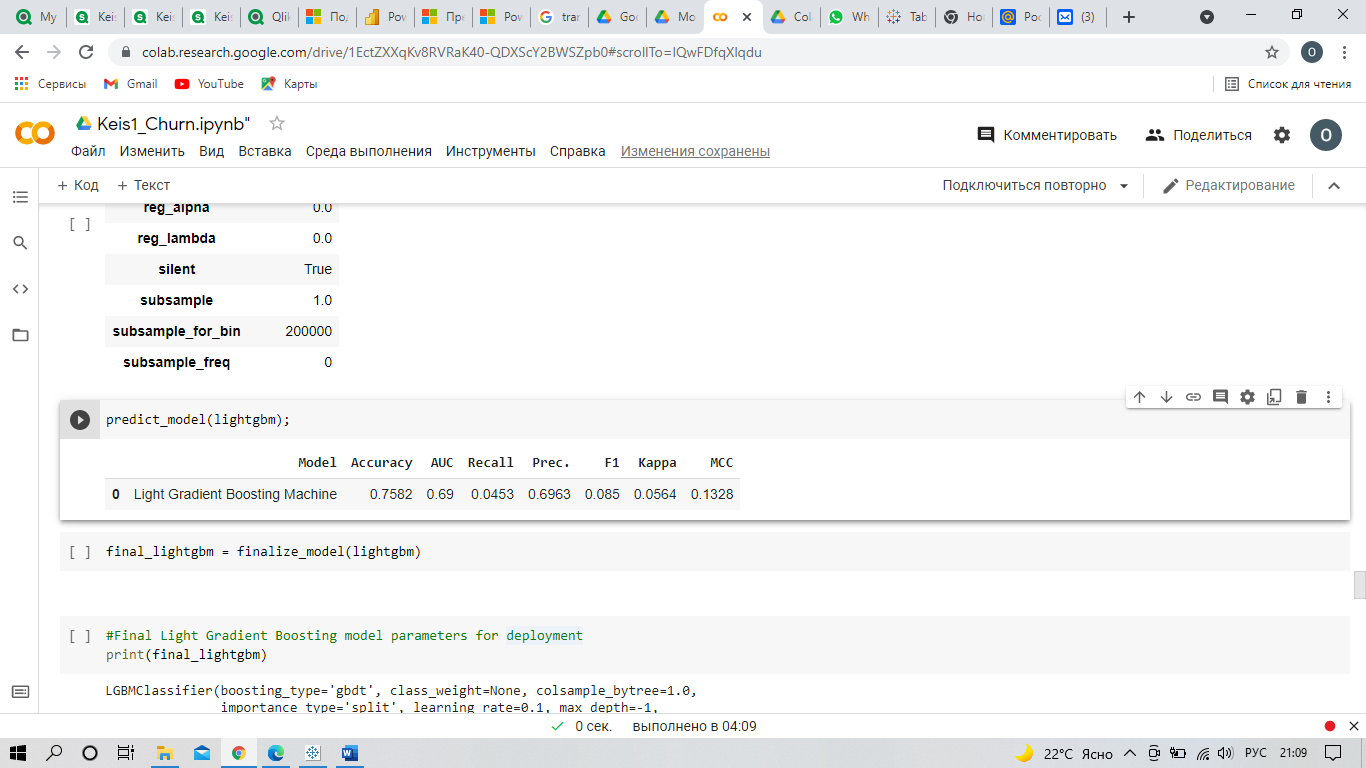
Далее в блокноте Colab было произведено EDA с использованием pandas-profiling и машинное обучение с использованием PyCaret

Для построения модели были выбраны следующие признаки: 'ABC\_XYZ','Region', 'Country', 'MaritalStatus', 'YearlyIncome', 'ProductName','ProductLine'.

Выходной столбец: 'Churn'.

По итогам машинного обучения лучшей моделью было выбрано Light Gradient Boosting.

Результаты:



В итоге результатами лучшего машинного обучения стала платформа RapidMiner.

Далее в BigQuery, Tableau Dekstop, Qlick Sense были созданы выражения KPI, построены дашборды и выявлены инсайты. Интерактивные отчеты были внедрены в Colab.

Ссылка Colab: https://colab.research.google.com/drive/1EctZXXqKv8RVRaK40-QDXScY2BWSZpb0?usp=sharing