Прогнозирование оттока клиентов телекоммуникационной компании

Выполнил Гилёв Юрий, 8-910-632-2855, astropix@yandex.ru

Постановка задачи

Модель, позволяющая предсказать вероятность оттока клиентов на основе данных телекоммуникационной компании.

Файлы проекта

- churn.txt – исходные данные  
- churn\_csv – исходные данные с названием столбцов в UTF-8  
- churn\_process.py – python-скрипт, делающий прогноз методом «Случайного леса»  
- churn\_predict\_50.csv – результат работы скрипта, загружается для окончательного анализа в EXCEL  
- churn\_excel.xlsx – EXCEL файл для финального анализа, подключающий исходные данные и результат

Модель компании на основе входных данных

Штат  
поле: Штат

На основании исходных данных можно утверждать:

- компания работает в нескольких штатах

- оказывает услуги связи

Представительство компании

Техподдержка  
поле: СЦ

- оказывает услуги международной связи

Услуги

- оказывает услугу

Телекоммуникации

- работает техподдержка

VMail  
поля:   
-Услуга VMail  
-Количество VMail

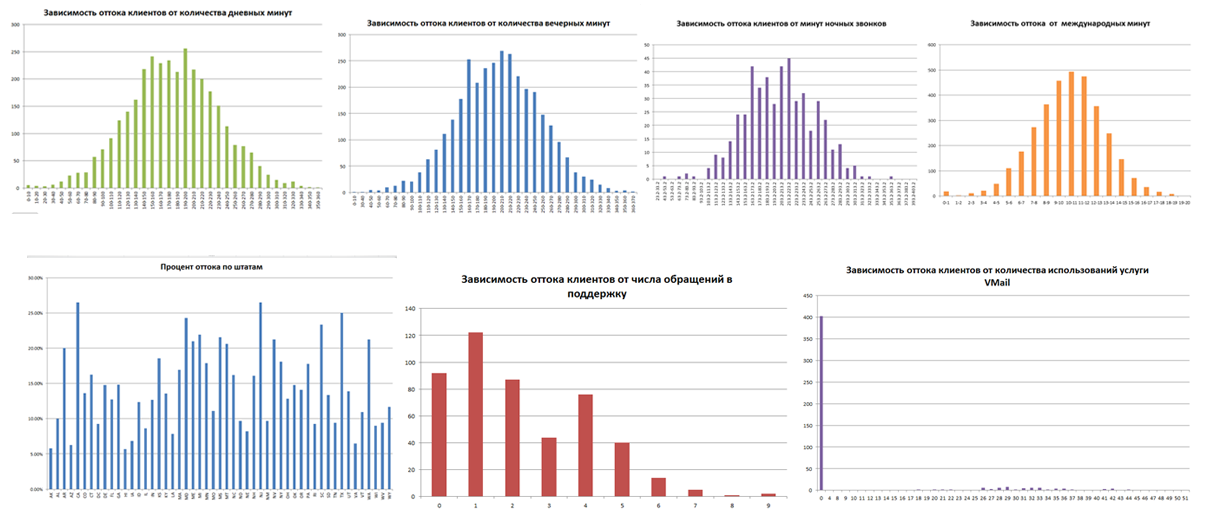
Ночные  
поля:   
-Ночные звонки  
-Ночная оплата  
-Ночные минуты

Вечерние  
поля:   
-Вечерние звонки  
-Вечерняя оплата  
-Вечерние минуты

Международные  
поля:   
-Межд. звонки  
-Межд. оплата  
-Межд. минуты

Дневные  
поля:   
-Дневные звонки  
-Дневная оплата  
-Дневные минуты

Знакомство с данными для дальнейшего анализа



Выводы

- отток клиентов по разным штатам варьируется от (5 до 25%). Следует проверить коррелирует ли это с открытием филиалов компании в разных штатах. Если нет, то проверить работу филиалов, конкурентную среду и бизнес-климат на местах.

- часть клиентов ушла, даже не обратившись в поддержку, что говорит о присутствии нелояльных клиентов (0.8% от общего числа)

- для лояльных клиентов критичным значением для оттока является 4 обращения в сервисный центр.

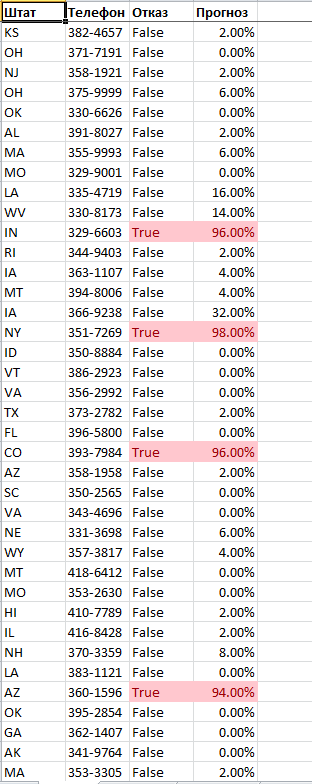
- услуга VMAIL не является ключевой для клиентов компании

- телекоммуникации это основная услуга компании. Клиенты компании примерно одинаково её используют в дневное, вечернее и ночное время. Кроме того, не обнаружен пик, связанный использованием корпоративных тарифов, кроме того, международная связь используется редко. Что говорит о том, что клиенты, это широкие слои населения.

- распределение вероятности оттока от числа минут хорошо коррелирует с нормальным распределением, что позволяет утверждать, что существует число минут, при котором отток клиента наиболее вероятен. Это 190,200,210 мин дневных, вечерних и ночных звонков, соответственно.

Модель поведения клиентов

На основании имеющихся данных можно утверждать, что подавляющее большинство клиентов лояльны компании и проходят все стадии жизненного цикла клиента: получение информации о компании, заключение договора, использование услуги, рост лояльности, уход. Разумеется, прогноз вероятности оттока следует использовать для проведения мероприятий по удержанию клиента.

Решение

Для решения задачи необходимо использовать методы, учитывающие следующие факторы: интенсивность использования услуг и различных тарифов компании и обращения в сервисный центр. Для прогнозирования вероятности оттока клиента предлагается использовать алгоритм машинного обучения «случайный лес». Обучить его на исходных данных и чтобы в дальнейшем использовать для прогноза.   
Был составлен Python-скрипт **churn\_process.py**, результаты работы которого загружаются в файл **churn\_excel.xlsx** Алгоритм показал высокую точность прогноза при числе деревьев n\_estimators=50.

Данное решение не учитывает возможные риски, и будет стабильным в случае постоянного бизнес климата. Кстати, сама компания довольно уязвима перед непредвиденными обстоятельствами, будь то: развитие новых технологий, стихийные бедствия, политическая нестабильность, биржевые спекуляции и т.д.