

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

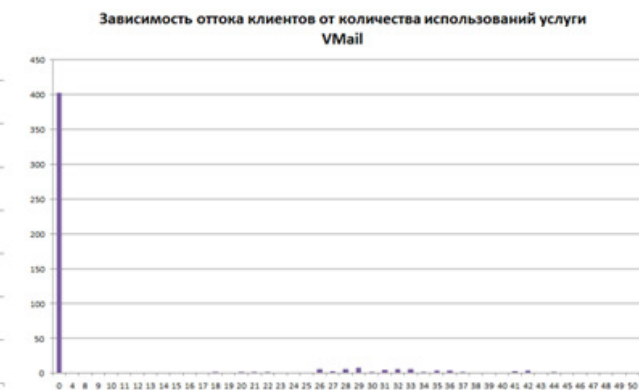
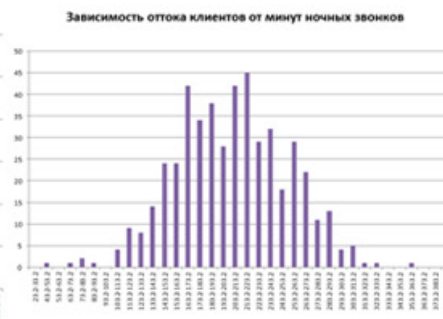
Модель компании на основе входных данных

На основании исходных данных можно утверждать:

- компания работает в нескольких штатах
- оказывает услуги связи
- оказывает услуги международной связи
- оказывает услугу
- работает техподдержка



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.

Штат	Телефон	Отказ	Прогноз
KS	382-4657	False	2.00%
OH	371-7191	False	0.00%
NJ	358-1921	False	2.00%
OH	375-9999	False	6.00%
OK	330-6626	False	0.00%
AL	391-8027	False	2.00%
MA	355-9993	False	6.00%
MO	329-9001	False	0.00%
LA	335-4719	False	16.00%
WV	330-8173	False	14.00%
IN	329-6603	True	96.00%
RI	344-9403	False	2.00%
IA	363-1107	False	4.00%
MT	394-8006	False	4.00%
IA	366-9238	False	32.00%
NY	351-7269	True	98.00%
ID	350-8884	False	0.00%
VT	386-2923	False	0.00%
VA	356-2992	False	0.00%
TX	373-2782	False	2.00%
FL	396-5800	False	0.00%
CO	393-7984	True	96.00%
AZ	358-1958	False	2.00%
SC	350-2565	False	0.00%
VA	343-4696	False	0.00%
NE	331-3698	False	6.00%
WY	357-3817	False	4.00%
MT	418-6412	False	0.00%
MO	353-2630	False	0.00%
HI	410-7789	False	2.00%
IL	416-8428	False	2.00%
NH	370-3359	False	8.00%
LA	383-1121	False	0.00%
AZ	360-1596	True	94.00%
OK	395-2854	False	0.00%
GA	362-1407	False	0.00%
AK	341-9764	False	0.00%
MA	353-3305	False	2.00%