**Описание проекта: телекоммуникации**

Оператор связи «ТелеДом» хочет бороться с оттоком клиентов. Для этого его сотрудники начнут предлагать промокоды и специальные условия всем, кто планирует отказаться от услуг связи. Чтобы заранее находить таких пользователей, «ТелеДому» нужна модель, которая будет предсказывать, разорвёт ли абонент договор. Команда оператора собрала персональные данные о некоторых клиентах, информацию об их тарифах и услугах. Ваша задача — обучить на этих данных модель для прогноза оттока клиентов.

**Описание услуг**

Оператор предоставляет два основных типа услуг:

* Стационарную телефонную связь. Телефон можно подключить к нескольким линиям одновременно.
* Интернет. Подключение может быть двух типов: через телефонную линию (DSL, от англ. digital subscriber line — «цифровая абонентская линия») или оптоволоконный кабель (Fiber optic).

Также можно подключить определённые дополнительные опции.

Клиенты могут платить за услуги каждый месяц или заключить договор на 1–2 года. Возможно оплатить счёт разными способами, а также получить электронный чек.

**Описание данных**

Данные состоят из нескольких файлов, полученных из разных источников:

* contract\_new.csv — информация о договоре;
* personal\_new.csv — персональные данные клиента;
* internet\_new.csv — информация об интернет-услугах;
* phone\_new.csv — информация об услугах телефонии.

**Файл contract\_new.csv**

* customerID — идентификатор абонента;
* BeginDate — дата начала действия договора;
* EndDate — дата окончания действия договора;
* Type — тип оплаты: раз в год-два или ежемесячно;
* PaperlessBilling — электронный расчётный лист;
* PaymentMethod — тип платежа;
* MonthlyCharges — расходы за месяц;
* TotalCharges — общие расходы абонента.

**Файл personal\_new.csv**

* customerID — идентификатор пользователя;
* gender — пол;
* SeniorCitizen — является ли абонент пенсионером;
* Partner — есть ли у абонента супруг или супруга;
* Dependents — есть ли у абонента дети.

**Файл internet\_new.csv**

* customerID — идентификатор пользователя;
* InternetService — тип подключения;
* OnlineSecurity — блокировка опасных сайтов;
* OnlineBackup — облачное хранилище файлов для резервного копирования данных;
* DeviceProtection — антивирус;
* TechSupport — выделенная линия технической поддержки;
* StreamingTV — стриминговое телевидение;
* StreamingMovies — каталог фильмов.

**Файл phone\_new.csv**

* customerID — идентификатор пользователя;
* MultipleLines — подключение телефона к нескольким линиям одновременно.

Во всех файлах столбец customerID содержит код клиента. Информация о договорах актуальна на 1 февраля 2020 года.

**Наш план:**

**1. Загрузка данных**

**2. Проведение исследовательского анализа и преобработки данных**

**3. Объединение данных в один большой датафрейм**

**4. Исследовательский анализ и преобработка данных большого (объединённого) датафрейма.**

**5. Подготовка данных для обучения моделей**

**6. Проведём обучение моделей**

**7. Выберем лучшую модель и проверим её качество на тестовой выборке, а также проверим качество по сравнению с dummy-моделью.**

**8. Напишем отчёт и рекомендации заказчику.**

В этом проекте будут использованы следующие библиотеки:

pandas, numpy, seaborn, scikit-learn (sklearn), lightgbm, catboost, phik.