**Национальный исследовательский университет**

**«Высшая школа экономики».**

Дисциплина «Интеллектуальные информационные системы».

Направление«Бизнес-Информатика» подготовки бакалавра.

Исследовательская работы на тему:

Churn Prediction for KKBOX service

Выполнили студенты:

Согоян Гурген Арамович, ББИ156

Римский Никита Валерьевич, ББИ153

Нестерова Милана Вадимовна, ББИ156

Преподаватель, проверяющий работу:

Шевгунов Тимофей Яковлевич

Москва, 2017

# Введение

Задача классификации это одна из классических задач, которую решают исследователи в области анализа данных. Понять к какому классу следует отнести тот или иной объект в зависимости от значений его параметров и быть максимально точным в этом предсказании это цель которую ставит для себя каждый data scientist, решая подобную задачу.

Стоит обратить внимание, что предсказывать класс объекта можно абсолютно в разных областях, начиная с одной из наиболее известной задачи классификации ирисов, заканчивая сложными моделями кредитного скоринга, дающую нам информацию о том, сможет ли клиент вернуть кредит или нет. В последние годы широкое распространение получил такой вид задач, как churn prediction, цель которого предсказать отток клиентов или иначе определить будет ли пользователь продолжать пользоваться сервисом компании или нет.

Очевидно, что любая компания заинтересована в том, чтобы человек продолжал пользоваться предоставляемыми ею услугами, а не отказывался, считая это бесполезным или считая, что уровень обслуживания достаточно низок, чтобы уйти к конкурентам. Если компании удастся вовремя определить подобную угрозу, то у нее еще есть шанс и время до конца подписки доказать клиенту, насколько важно и полезно то, что для него делает компания.

Одной из таких компаний является KKBOX, крупнейший азиатский музыкальный сервис, предоставляющий безграничный доступ к библиотеке, содержащей более 30 миллионов композиций, в том числе малоизвестных азиатских артистов. Соответственно, в зависимости от срока подписки музыке клиент должен будет заплатить определенную сумму. Через небезызвестный сайт kaggle.com в наши руки попала база данных подписчиков данного сервиса, информации о них, об их действиях об их транзакциях и т.д., а также информация о том, остались они клиентами или нет.

**Объектом исследования** работы является база данных клиентов KKBOX

**Предметом исследования** является модель, предсказывающая продолжение подписки пользователя.

Таким образом, целью нашего исследование является построение модели, предсказывающей продолжен ли клиент пользоваться услугами KKBOX, или перестанет быть ее клиентом, определение факторов, влияющих на значение целевой переменной, а также поиск модели, дающей наименьшую ошибку при тестировании данных. Для достижения цели нам будет необходимо решить следующие задачи: