

Оптимизация предсказания оттока абонентов оператора сотовой связи

Выполнила: Сулягина Анастасия 371 гр.

Научный руководитель: Невоструев Константин, ведущий разработчик ООО «НМТ»

Введение

- Удержание абонента дешевле привлечения нового
- Отток ~25% в год
- Машинное обучение позволяет достаточно точно предсказать поведение абонента

Задача

Оптимизировать предсказание оттока абонентов с мобильного оператора, полученное в прошлом году, используя методы машинного обучения.

- Провести сравнительный анализ существующих классификаторов
- Провести настройку моделей для улучшения результатов
- Сгруппировать данные для получения более точных результатов

Обзор

Лучший результат прошлого года: Precision: 0.75, Recall: 0.66, AUC: 0.9

Наиболее перспективные модели

- Градиентный бустинг
- Нейронная сеть
- Ансамбли классификаторов

Данные

- 50000 пользователей
- Активность использования различных видов связи
- Промежуток активности 15 месяцев
- Персональные данные

Классификация

Модель	precision	recall	AUC	accuracy
Ансамбль*	0.75	0.72	0.92	0.88
Градиентный бустинг	0.74	0.69	0.92	0.87
Нейронная сеть	0.70	0.62	0.89	0.86
Ближайшие соседи + лес	0.74	0.60	0.86	0.87

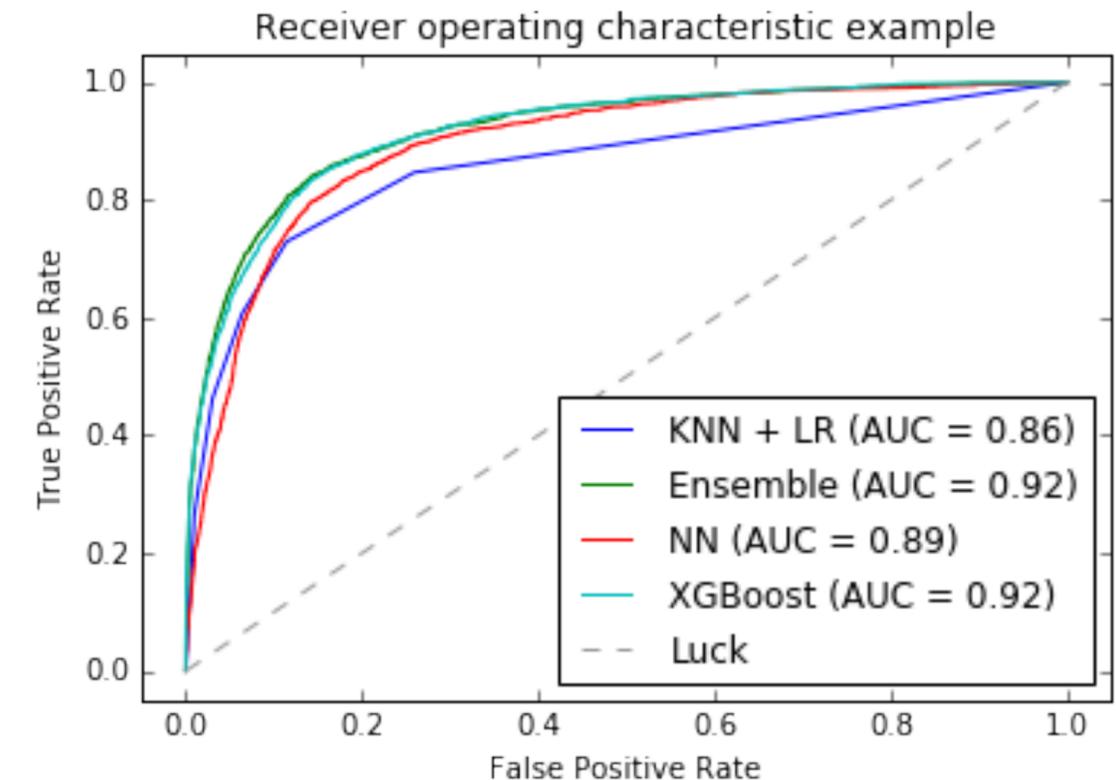
precision - действительно уйдут из всех угаданных нами

recall - угаданы нами из всех, кто уйдет

accuracy - частота корректных предсказаний

ROC-кривая - соотношение между угаданными ушедшими и неверно угаданными ушедшими

*Состоит из двух RandomForest, двух ExtraTrees с разными функциями качества и одного бустинга



Настройка

Модель

Параметры

Ансамбль

Тип классификаторов, способ их взаимодействия

Градиентный бустинг

Количество деревьев, количество признаков для обучения

Нейронная сеть

Тип слоя, количество слоев, их порядок и размер

Группировка

Модель	precision	recall	AUC	accuracy
юр.лицо	0.81	0.93	0.87	0.81
физ.лицо	0.70	0.59	0.90	0.90
> 2 лет с оператором	0.75	0.58	0.86	0.94
< 2 лет с оператором	0.74	0.79	0.86	0.78
< 40 лет	0.71	0.63	0.86	0.90
> 40 лет	0.79	0.77	0.93	0.89
пользуются интернетом	0.74	0.73	0.92	0.88
пользуются только связью	0.75	0.69	0.91	0.89

Результаты

- Улучшены результаты прошлого года

	precision	recall	AUC
Моя работа	0.75	0.72	0.92
Работа прошлого года	0.75	0.66	0.90

- Реализованы классификаторы, настроены их параметры
- Проведен сравнительный анализ реализованных классификаторов
- Проведена классификация на сгруппированных данных