

Использование линейного  
дискриминантного анализа в  
задаче предсказания оттока  
абонентов оператора сотовой  
СВЯЗИ

Выполнил: Минаев Александр, 371 гр.  
Научный руководитель: д.ф.-м.н., проф. А.Н. Терехов

# Введение

- Отток абонентов может достигать отметки в 50%
- Привлечение новых абонентов требует больше денег и ресурсов, чем удержание старых
- Задача предсказания оттока абонента – это задача бинарной классификации
- С помощью машинного обучения можно получить хорошую точность предсказания

# Цель работы

Реализация классификатора на основе линейного дискриминантного анализа, предсказывающего отток абонентов.

Задачи:

- Проанализировать существующие решения
- Провести предобработку данных
- Разработать модель
- Оптимизировать результаты

# Метрики

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$TPR = \frac{TP}{TP + FN}$$

$$recall = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

# Обзор

Лучший результат предыдущих лет с помощью XGBoost:

AUC	Recall	Precision
0.92	0.72	0.75

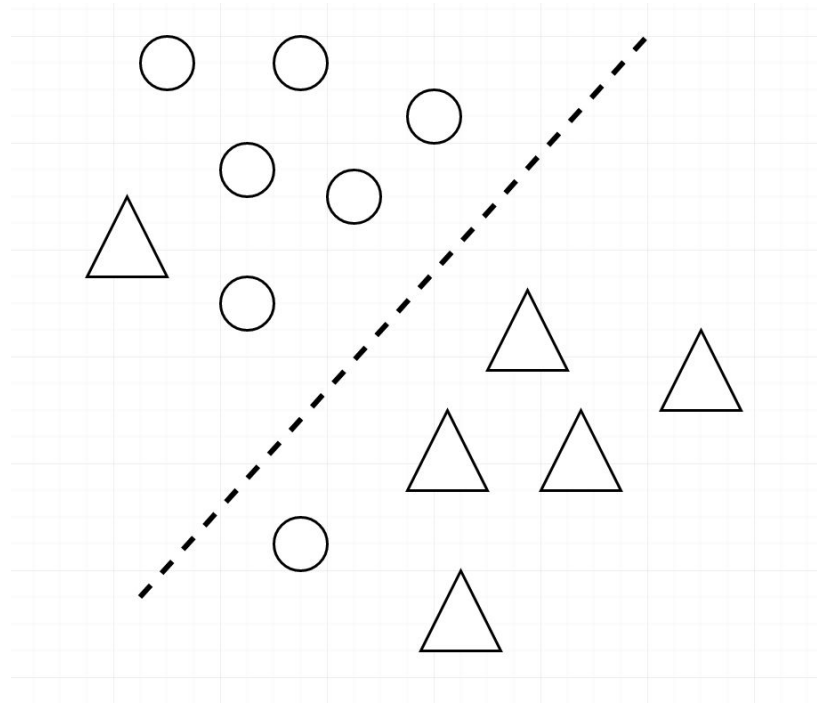
LDA показывает не лучшие результаты, однако бэггинг над ним выглядит перспективно.

# Данные

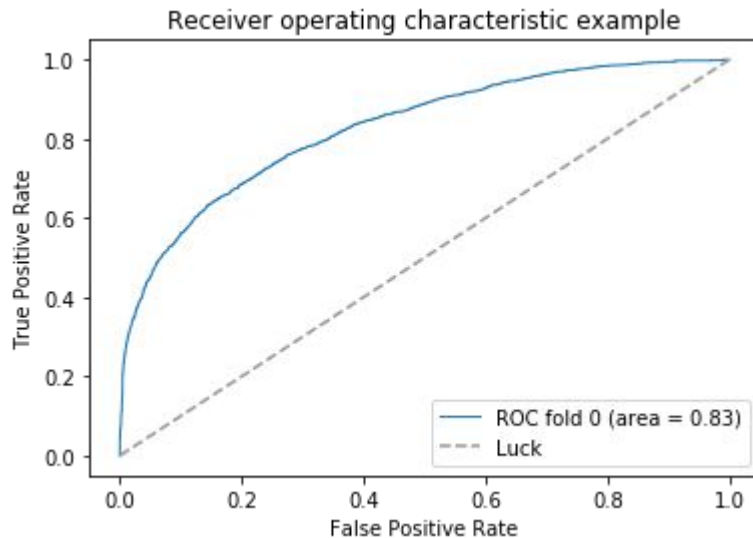
Данные были агрегированы по следующим категориям:

- Минуты и стоимость входящих вызовов
- Минуты и стоимость исходящих вызовов
- Количество и стоимость исходящих СМС
- Количество входящих СМС
- Количество трафика мобильного интернета и его стоимость
- Информация об обращениях клиента в службу поддержки
- Тарифный план клиента
- Личные данные

# Линейный дискриминантный анализ



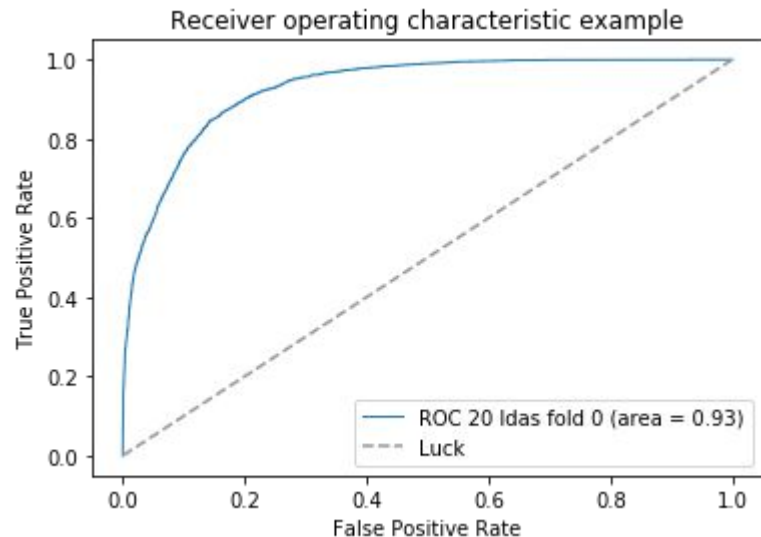
# Результаты LDA



AUC	Recall	Precision
0.83	0.58	0.54



# Результаты бэггинга над LDA



AUC	Recall	Precision
0.93	0.75	0.67

# Результаты

- Проанализированы существующие решения
- Проведена обработка данных
- Реализована модель LDA
- Улучшены результаты с помощью построение бэггинга над LDA
- Улучшены результаты предыдущих лет

	AUC	Recall	Precision
Лучший результат предыдущих лет	0.92	0.72	0.75
LDA	0.83	0.58	0.54
Бэггинг над LDA	0.93	0.85	0.58