

# Определение популярности геолокации для размещения банкомата

<https://github.com/AAVMLDIS/Project>

<https://project-1-nko7.onrender.com/docs>

Алексей Кирьянов  
Владимир Мирошников  
Александра Григорьева

### Источники данных

- Open Street Map
- mainfin.ru
- boosters.pro

### Индекс популярности геолокации банкомата (данные по банкоматам Росбанка)

-0.15 . . . 0.2

худшая оценка

лучшая оценка

### Данные

- **Города** (Москва, Санкт-Петербург, Екатеринбург, Новосибирск, Челябинск, Омск, Казань, Красноярск, Краснодар, Нижний Новгород, Уфа, Ростов-на-Дону, Пермь, Самара, Волгоград, Воронеж)
- **Объекты** (автобусные остановки, больницы, метро, жд станции, отели, супермаркеты, тц, магазины, аптеки, университеты, колледжи, стадионы, кафе, рестораны, фуд корты, заправки, обменники, кинотеатры, ночные клубы, театры, полицейские участки, суды)

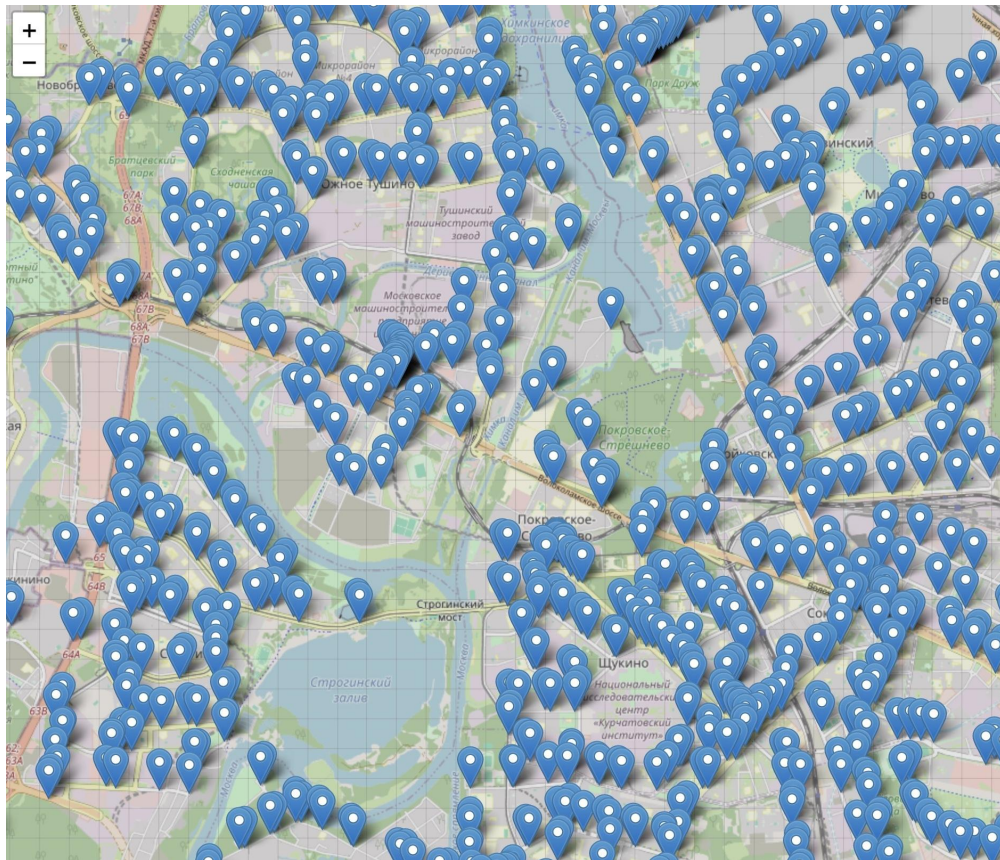
К решению задачи мы изначально подошли с построения сетки на карте и распределения различных объектов по квадратам.

### Построение сетки

Определили координаты северо-западной точки

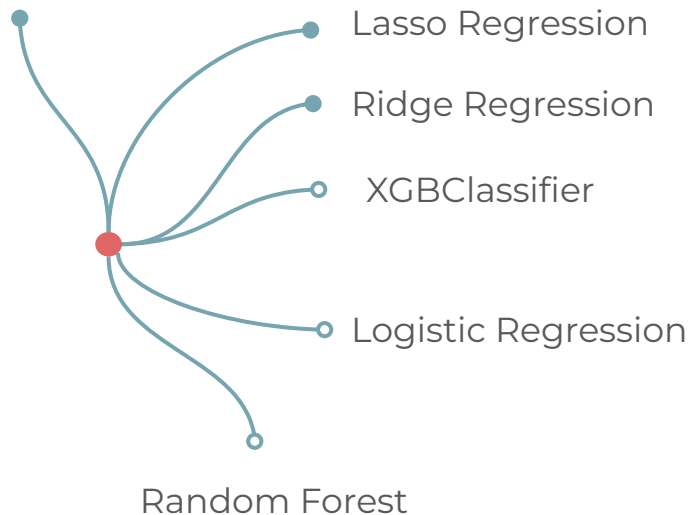
Определили протяженность города на восток и на юг

Отложили квадраты 300x300 по всей площади города



## Модели, которые пробовали

Random Forest Regression



## Фичи, которые использовали в моделях

Минимальное расстояние от центра ячейки до объекта

Количество объектов каждого типа в ячейке

Численность населения

Расстояние от центра ячейки до административного центра города

Таргет

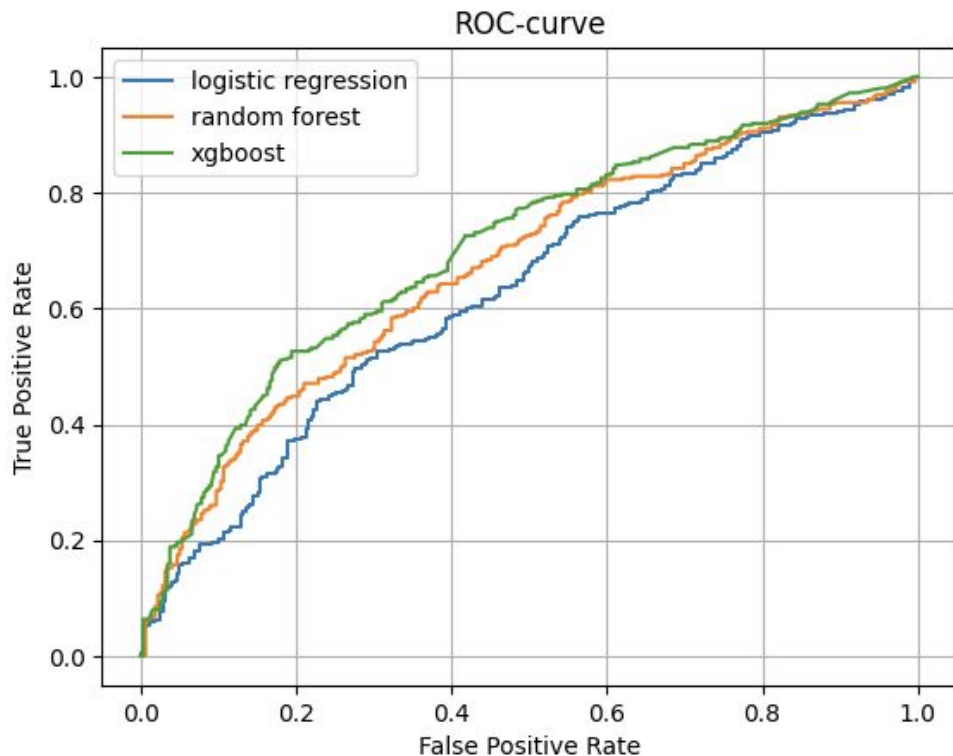
**бинарная  
классификация**

1 (score > 0)  
0 (score < 0)

**регрессия**

score

## Модели, на которых мы остановились



### Logistic Regression

Gini = 0.199

accuracy = 0.59

**AUC = 0.599**

### Random Forest

Gini = 0.244

accuracy = 0.62

**AUC = 0.621**

### XGBClassifier

Gini = 0.304

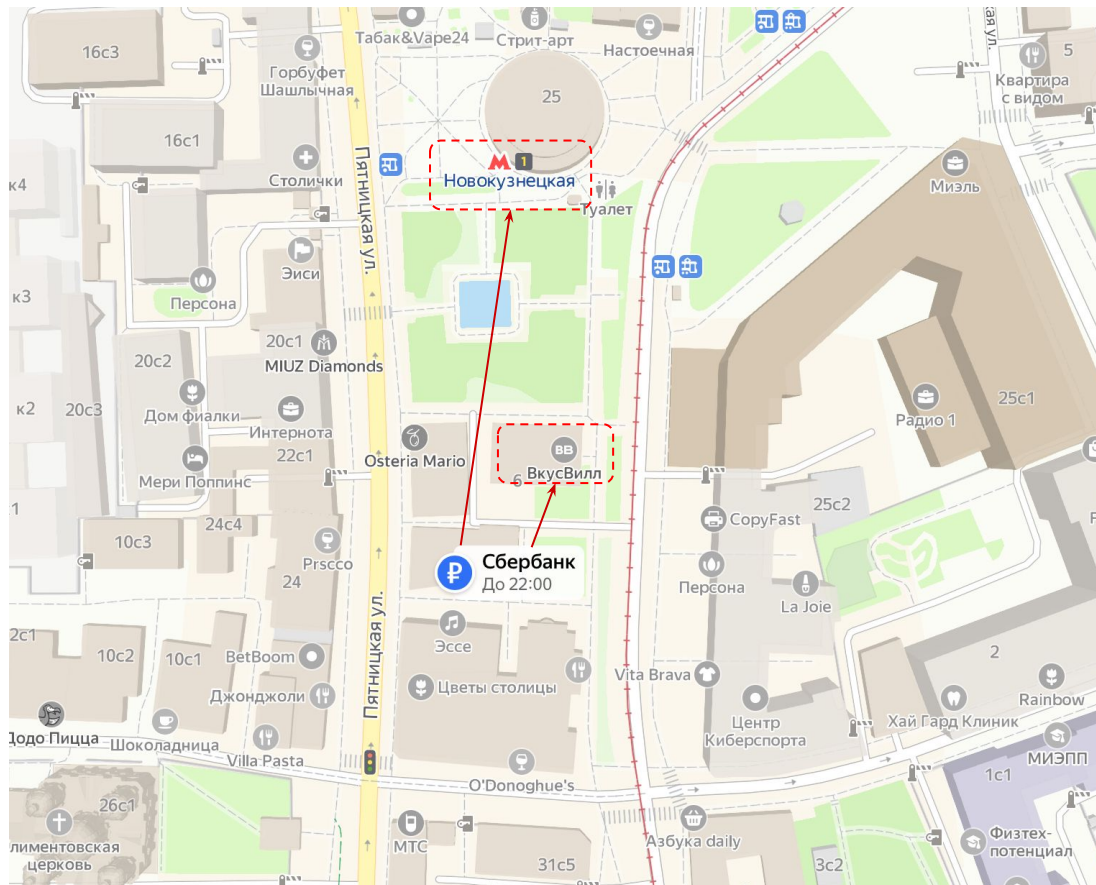
accuracy = 0.65

**AUC = 0.652**

В следующем шаге в качестве предикторов мы попробовали использовать **расстояние** до ближайших от банкоматов объектов. Например, расстояние от банкомата до ближайшей станции метро.

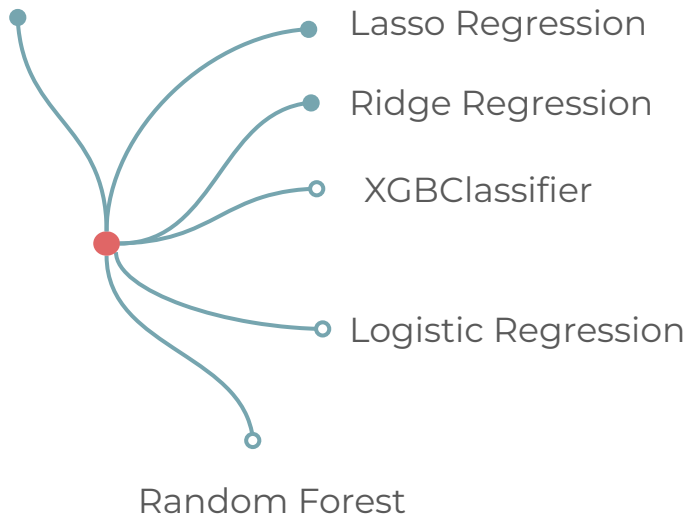
Также мы изменили фичи с прошлого чекпойнта (удалили фичи, связанные, с разделением карты на квадраты, а также добавили несколько новых)

- Количество объектов каждого типа в радиусе 100 и 300 метров
- Численность населения
- Плотность населения
- Прожиточный минимум
- Средняя заработная плата



## Модели, которые пробовали

Random Forest Regression



Таргет

бинарная  
классификация

1 (score > 0)  
0 (score < 0)

регрессия

score

## Модели, на которых мы остановились

**Random Forest**

$R^2 = 0.192$

GridSearch score = 0.183

**XGBClassifier**

GridSearch score = 0.165

API сервис (<https://api-atms-pcm9.onrender.com/docs#/>)

1



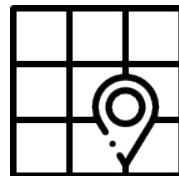
на вход сервис  
принимает адрес



данные передаются  
на DaData



адресу  
сопоставляются  
координаты на карте



по координатам  
определяется  
ячейка в сетке



на выходе сервис  
возвращает метки  
класса

2



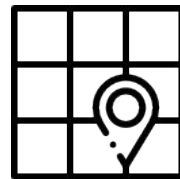
на вход сервис принимает  
csv с адресами



данные передаются  
на DaData



адресу  
сопоставляются  
координаты на карте



по координатам  
определяется  
ячейка в сетке



на выходе сервис  
возвращает csv с  
адресами и метками  
класса



## Что планируем сделать в дальнейшем?

Что уже успели сделать с прошлого чекпойнта?

- Добавили новые фичи
- Улучшили показатели по моделям
- Начали работу над тг ботом



Продолжить работу с фичами



Продолжить работу с моделями



Создать бота и сделать визуализацию в API