

# Прогнозирование показателей миграции в РФ с использованием моделей машинного обучения

Автор Семенов Роман Валерьевич





#### Цель и бизнес-задача

# Перед аналитическим отделом группы компаний работающей в сфере инвестиционной иммиграции в страны ЕС и Карибского бассейна, а также предоставляющей услуги по программе ЕВ-5 в США, в связи с запланированным открытием департамента релокации в Москве для работы с иностранными гражданами (экспатами), руководством поставлена задача спрогнозировать число прибывших на территорию РФ с 1 кв. 2021 года по 4 кв. 2021 года включительно.

# Настоящий анализ и его интерпретация будут включены в комплекс стратегического планирования и бюджетирования группы компаний.

# Проект призван оптимизировать расходы на департамент релокации и предстоящий комплекс маркетинга при выводе на рынок новой услуги.

# Постановка задачи машинного обучения



Визуализация корреляционной зависимости с помощью SeaBorn

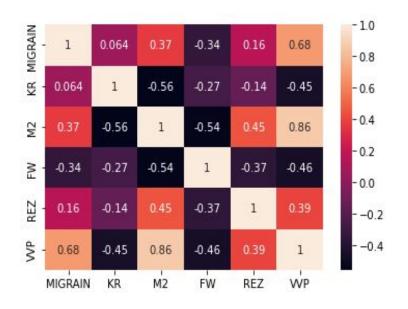
Построение диаграммы рассеивания с вектором найденных коэффициентов m и b в модели линейной регрессии

Прогнозирование с использованием моделей регрессии

- # линейная регрессия «LinearRegression»
- # метод ближайшего соседа «KneighborsRegressor»
- # дерево решений «DecisionTreeRegressor»
- # случайный лес «RandomForestRegressor»



## Набор данных (Dataset)



Пять сценарных (прогнозных) показателей ЦБРФ с 1 кв. 2021 года по 4 кв. 2021 года включительно и пять фактических показателей за период с 1 кв. 2013 года по 4 кв. 2020 года включительно

# КР - Ключевая ставка ЦБ РФ (% годовых)

# M2 - Темп прироста денежной массы в национальном распределении (млрд. руб.)

# FW - Цена на нефть марки Urals, средняя за год (долл. США за баррель)

# REZ - Изменение международных резервов РФ (% к предыдущему году)

# VVP - Валовый внутренний продукт (млрд. руб.)

Данные росстата по прибывшим с 1 кв. 2013 года по 4 кв. 2020 года включительно

# MIGRAIN - Число прибывших (человек)



### Модели



найденные коэффициенты:

сдвиг прямой - b = 48548.60979889624 наклон прямой - m = [4.35682775]



параметры:

n\_neighbors=5 (кол-во соседей)

Дерево решений

параметры:

min\_samples\_split = 3, min\_samples\_leaf = 3

#### Случайный лес

параметры:

min\_samples\_split = 3, min\_samples\_leaf = 3, n\_estimators = 16





## Результаты

«Линейная регрессия» LinearRegression		«Ближайший сосед» KNeighborsRegressor		«Дерево решений» DecisionTreeRegressor		«Случайный лес» RandomForestRegressor	
$R^2$	MAPE	$R^2$	MAPE	$R^2$	MAPE	$\mathbb{R}^2$	MAPE
0.502439341 1478226	0.115037368 84929057	0.495670713 56285186	0.110445765 76475315	0.72014860 41375744	0.170090103 6913446	0.64633758 52889883	0.08800950 391803786
Прогноз прибывших на период с 1 кв. 2021 г. по 4 кв. 2021 г. включительно							
174033.52703482		171278.2		171278.2		159937.59558532	
170268.79217415		171278.2		171278.2		153829.9609127	
174358.54638514		171278.2		171278.2		156214.62237103	
176886.37784696		171278.2		171278.2		156214.62237103	



### Выводы

02

«Дерево решений» DecisionTreeRegressor

лучший показатель **R^2=0.72** 

03

Внедрение модели

возможно

**О1 Корреляция**недостаточна



#### СПАСИБО ЗА ВНИМАНИЕ!