

# Прогноз эмиссии CO<sub>2</sub> автомашин



# Цель данной работы:

Главная цель - уменьшить количество выброс диоксида углерода и тем самым улучшить экологическое положение на нашей планете путем помощи автолюбителям совершать рациональный с точки зрения экологии выбор при покупке автомобиля.

Наша команда видит Задачу данного проекта-разработать модель, которая достаточно точно может предсказать количество CO<sub>2</sub> в выхлопе автомобиля, основываясь на различных технических характеристиках автомобиля.

Основными инструмента для решения поставленной задачи является язык программирования python и модули: NumPy, Matplotlib, SciPy, scikit-learn, and Streamlit

В качестве инструмента предсказания мы использовали машинное обучение, а в частности модели: Linear Regression, Random Forest, K-Nearest Neighbors (KNN), и Support Vector Regression (SVR)

# Наши Данные:

Для данной работы мы использовали данные об эмиссии диоксида углерода машин в Канаде, с 2000 по 2014 год



Government  
of Canada

Gouvernement  
du Canada



2010-2014 Fuel Consumption Ratings (2020-03-17)

English

dataset

CSV

Created	March 31, 2017
Format	CSV

## Структура данных

- Make: Бренд автомобиля.
- Model: конкретная модель автомобиля.
- Vehicle\_class: тип кузова автомобиля
- Engine\_size: объем двигателя автомобиля в литрах.
- Cylinders: количество цилиндров.
- Transmission: тип коробки передач (например, автоматическая, ручная).
- Fuel\_type: тип топлива, используемого автомобилем.
- Fuel\_consumption\_city: показатели расхода топлива в городе в литрах на 100 километров.
- Fuel\_consumption\_hwy: показатели расхода топлива на шоссе в литрах на 100 километров.
- Fuel\_consumption\_comb(l/100km): рейтинг комбинированного расхода топлива (город и трасса) в л/100 км.
- Fuel\_consumption\_comb(mpg): общий рейтинг расхода топлива в милях на галлон (миль на галлон).
- Co2\_emissions: выбросы углекислого газа в выхлопных трубах при комбинированном движении по городу и шоссе, в граммах на километр.

## Подготовка данных:

1. Мы переформатировали столбцы дав им более удобные названия изменив регистры
2. Избавились от ненужных столбцов которые не несли пользы для нашей модели(марка машины)
3. Избавились от дубликатов и пропусков
4. Привели к числовым типам данным некоторые столбцы в частности тип топлива, вид трансмиссии
5. Аргументировали данные и избавились от выбросов

# Анализ данных: Разведывательный анализ

vehicle class ↕	proportion ↕
SUV	17.331566
COMPACT	15.776787
MID-SIZE	13.860431
PICKUP TRUCK - STANDARD	11.715078
SUBCOMPACT	9.786670
TWO-SEATER	5.122333
FULL-SIZE	5.098228
STATION WAGON - SMALL	4.483548
MINICOMPACT	3.398819
STATION WAGON - MID-SIZE	3.025190

Распределение типов машин в  
канаде.

engine size (l) ↕	proportion ↕
2.0	9.208148
3.0	8.545257
2.5	7.448475
3.5	7.291792
2.4	6.291431
5.3	3.495239
3.6	3.471134
1.8	3.254188
4.0	3.242136
1.6	2.651561

Распределение машин  
по объему двигателя

# Анализ данных: Разведывательный анализ

fuel type ↕	proportion ↕
X	54.48
Z	40.41
E	3.93
D	0.94
N	0.24

Распределение машин по топливу, где:

X - Обычный бензин;

Z - Бензин премиум-класса;

D - Дизельное топливо;

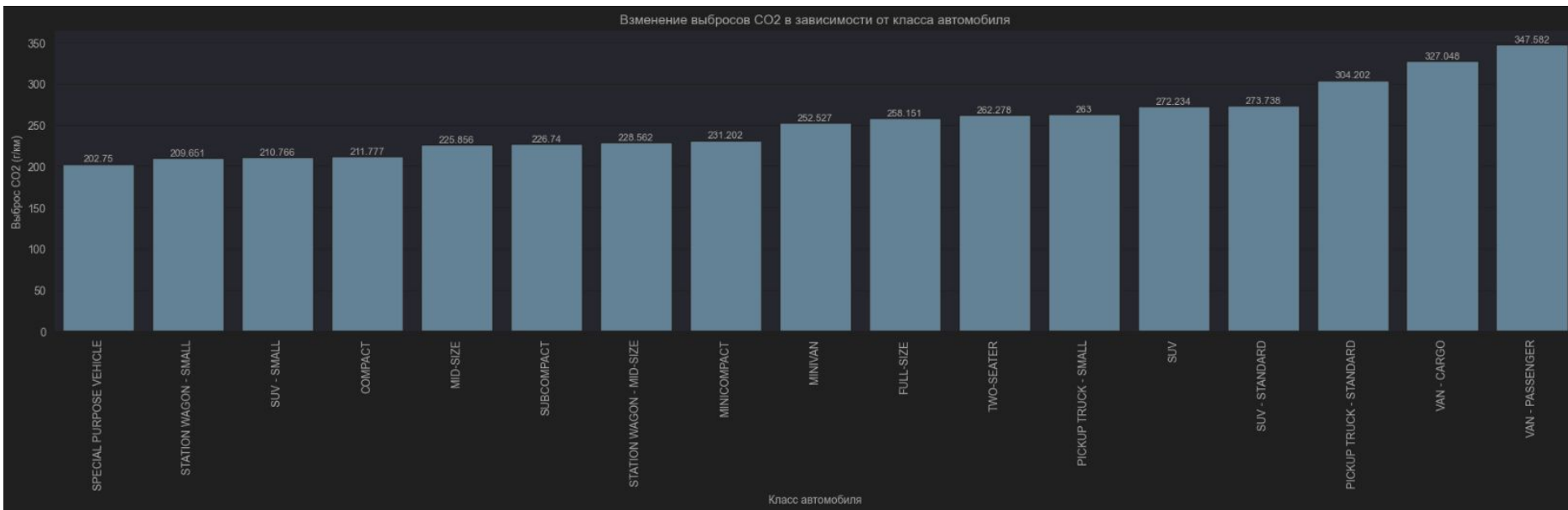
E - Этанол;

N - Природный газ.

transmission ↕	proportion ↕
Автомат	45.0
Ручная	26.0
Автоматическая с выбранным сдвигом	24.0
Бесступенчатая регулировка	3.0
Автоматическая ручная	1.0

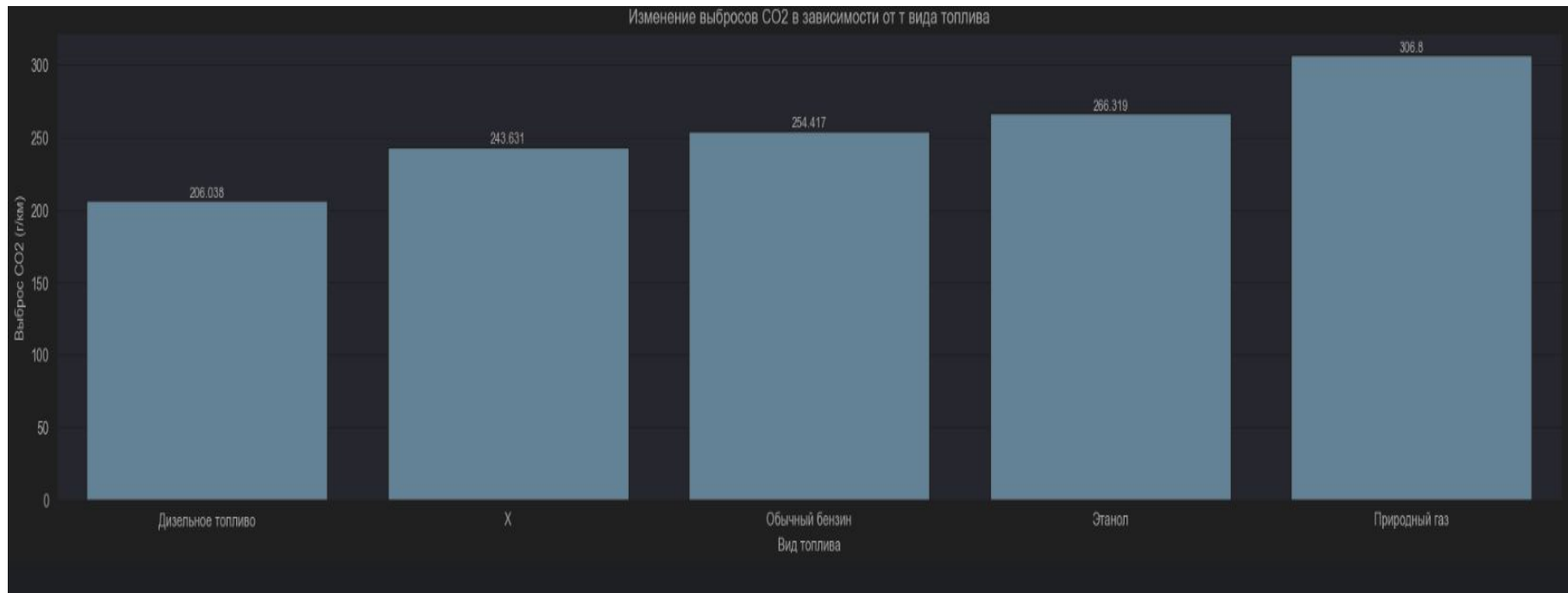
Пропорция трансмиссий в машинах в канаде

# Первые выводы по датасету:



Больше всего CO2 выделяют пассажирские автобусы и грузовые автомобили.

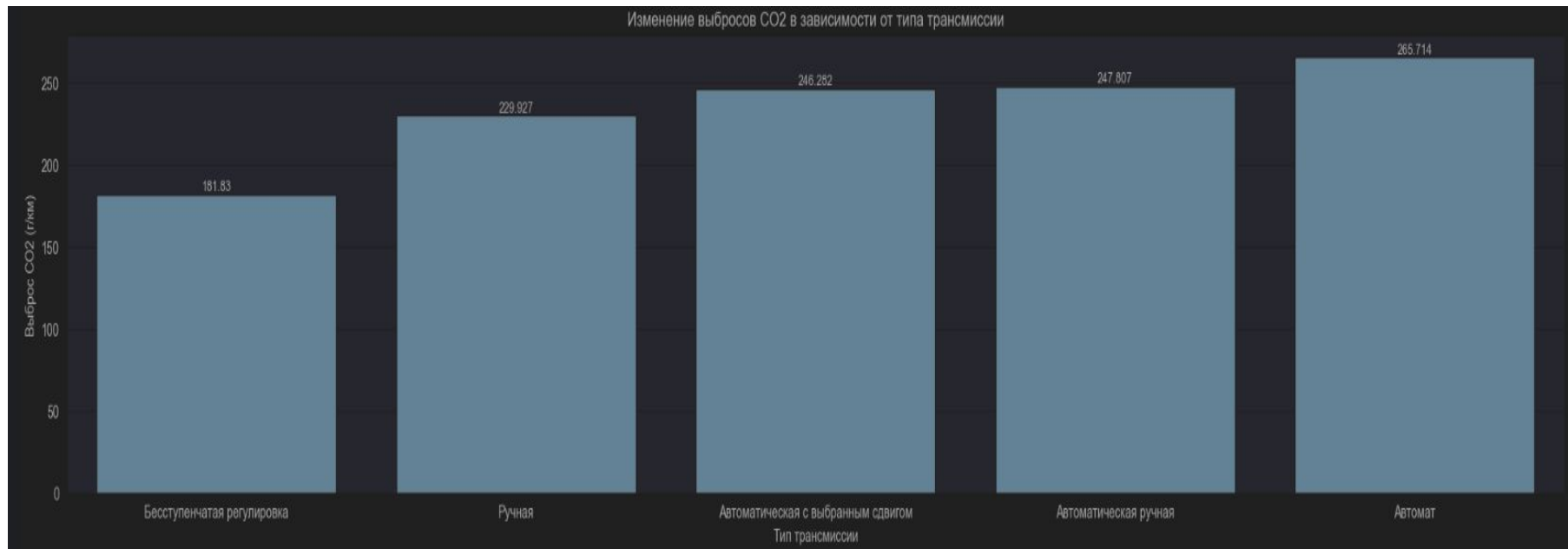
# Первые выводы по датасету:



В среднем меньше всего выбросов от дизельного топлива больше всего от природного газ.



# Первые выводы по датасету:



Машины с автоматической трансмиссией в среднем приносят больше CO2 в атмосферу

# Выбор модели:

↕	Model	↕	R2 Score ↕
0	Linear Regression		0.923856
1	Random Forest		0.956260
2	SVR		0.924719
3	KNN		0.957878

Для решения поставленной задачи мы решили сравнить эффективность четырех моделей

Как видно из таблицы, более высокую точность показал метод К-ближайших соседей (KNN). Его мы и будем использовать как основной метод предсказания.

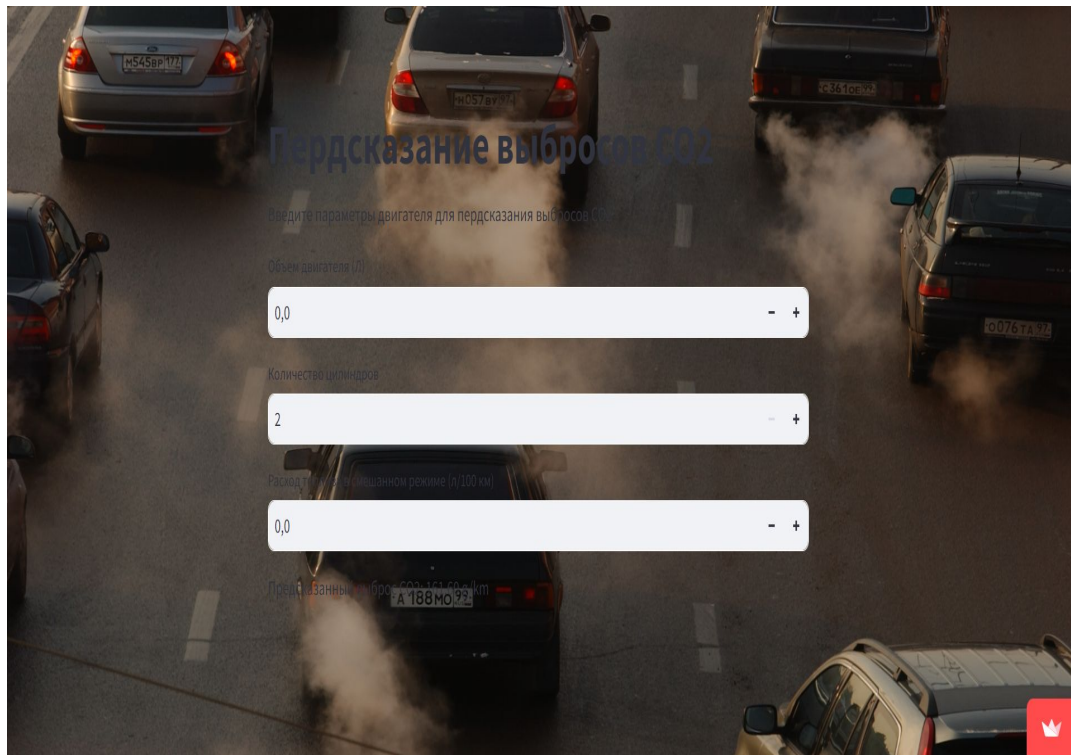
# Сводная таблица результатов моделей

Values from our Data Seta	Linear Model Predicated Values	KNN Model Predicated Values	SVR Model Predicated Values	RF Model Predicated Values
340	329.37	339.8	333.12	340
292	243.36	263.2	246.32	253.54
246	245.73	246	246.26	246
334	321.67	252.6	325.29	269.4
221	222.98	220	220.09	221
182	182.86	181.8	181.42	181.95
214	212.81	212.8	213.23	214
218	216.07	218.4	217.08	216.16
245	334.74	242.4	329.45	242.56
235	228.94	232	233.57	232
198	197.89	198	197.26	198

# Итог работы:

Итогом нашей работы стал калькулятор который позволяет рассчитать выбросы CO2 в зависимости от характеристик машины.

<https://co2-cars-emission-lekduu3e9dxuffsfwnxax9.streamlit.app/>



# Над проектом работали:

## Команда 15:

Азмукова Вероника

Абдрахманов Антон

Алексеев Николай

Базавлук Владимир

**Спасибо за внимание!**