



Дипломный проект на тему:

**«ПРЕДСКАЗАНИЕ СОЦИАЛЬНО-
ЭКОНОМИЧЕСКОГО РАЗВИТИЯ
РЕГИОНОВ РОССИИ»**

Слушатель:

Казанцев Егор Александрович

АКТУАЛЬНОСТЬ ТЕМЫ И ЕЕ ПРОБЛЕМАТИКА

ЦЕЛЬ ПРОЕКТА:

Создать цифровую рабочую модель предсказания (прогнозирования) уровня социально-экономического развития регионов России с применением методов машинного обучения

ЗАДАЧИ РАБОТЫ:

1. Изучить предметную область
2. Сформировать и разработать архитектуру данных
3. Собрать данные, провести их анализ
4. Подобрать и применить алгоритмы машинного обучения
5. Получить предсказание и интерпретировать его

АКТУАЛЬНОСТЬ И ПРОБЛЕМАТИКА:

Внедрение современных цифровых моделей и алгоритмов в традиционную деятельность, на примере сферы экономического прогнозирования (планирования), цифровая трансформация деятельности

ПРЕДСКАЗАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ РЕГИОНОВ РОССИИ

Этап 1: изучение предмета

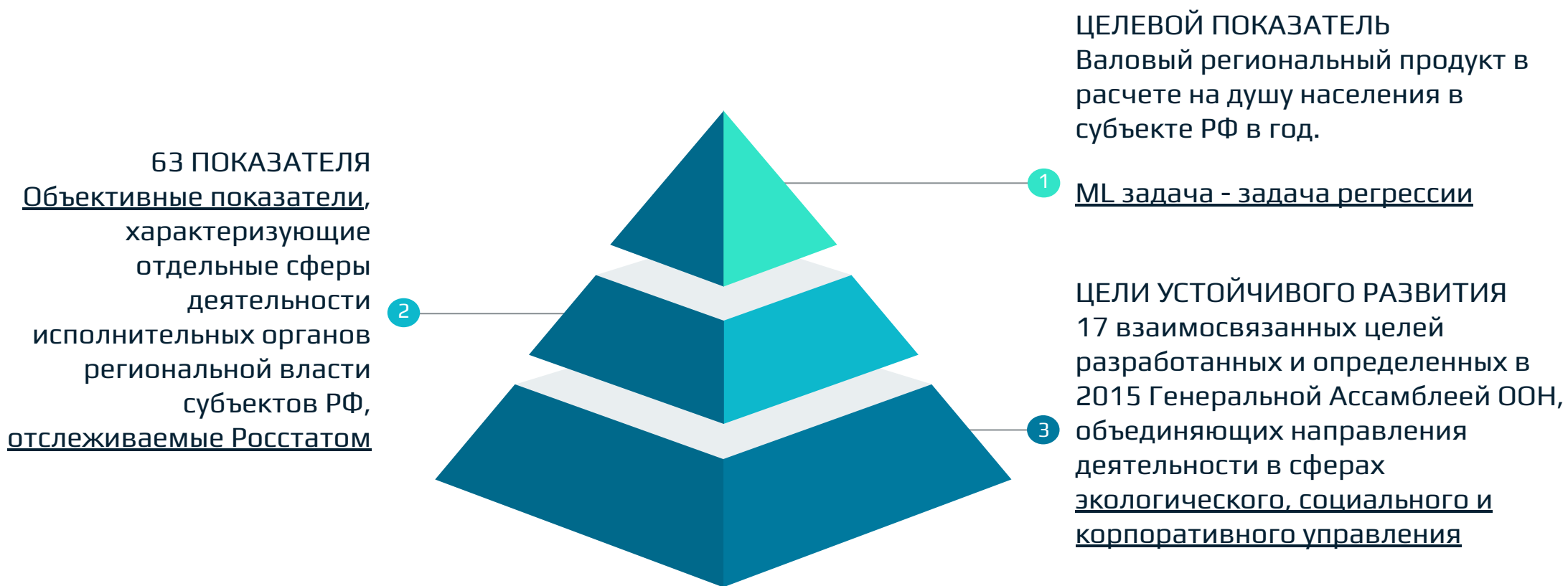


Рис. 1 Схематичное изображения предмета проекта

ПРЕДСКАЗАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ РЕГИОНОВ РОССИИ

Этап 2: поиск и сбор данных

ТРЕБОВАНИЯ К ДАННЫМ

1. Период выборки данных с 2015 по 2020 год
2. Данные региональной статистики по регионам России (субъектам Российской Федерации)
3. Числовые данные

ИСТОЧНИКИ ДАННЫХ

1. Статистический ежегодник Росстата Цели устойчивого развития в Российской Федерации, 2021 год
2. Данные, размещенные на сайте Росстата (<https://rosstat.gov.ru>)
3. Единая межведомственная информационно-статистическая система (ЕМИСС) (<https://www.fedstat.ru/>)

ПРЕДСКАЗАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ РЕГИОНОВ РОССИИ

Этап 3: анализ и структурирование

	Признак 1 за ... год	...	Признак 63 за ... год
Субъект РФ 1	Float64 / int64	Float64 / int64	Float64 / int64
Субъект РФ 2	Float64 / int64	Float64 / int64	Float64 / int64
Субъект РФ 3	Float64 / int64	Float64 / int64	Float64 / int64
....	Float64 / int64	Float64 / int64	Float64 / int64
Субъект РФ 85	Float64 / int64	Float64 / int64	Float64 / int64

РЕЗУЛЬТАТ:

- сформировано **17** таблиц данных, содержащих **63** признака, по каждому из **85** субъектов за 2015-2020 годы (всего **378** признаков)
- подготовлены обучающая (0,8) и тестовая (0,2) выборки с признаками за 2015-2017 годы, выборка для предсказания за 2018-2020 годы

ПРЕДСКАЗАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ РЕГИОНОВ РОССИИ

Этап 3: анализ и структурирование

ПРИЗНАКИ ОБЪЕКТОВ:

Значительной взаимосвязи
между признаками и целевой
переменной не выявлено

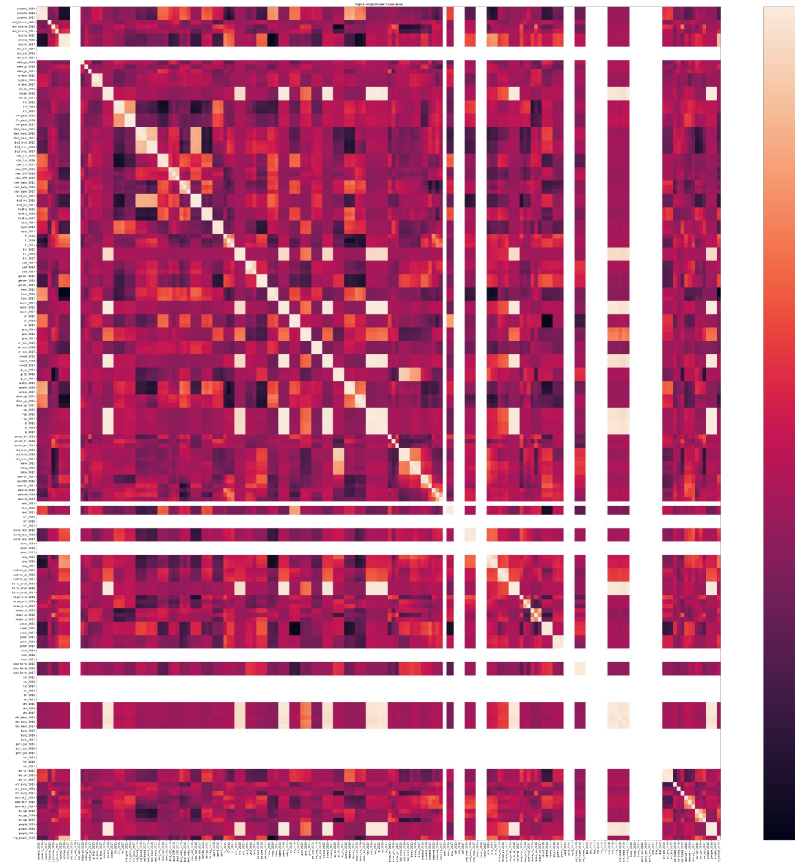


Рис. 2 Диаграмма корреляции признаков

ПРЕДСКАЗАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ РЕГИОНОВ РОССИИ

Этап 3: анализ и структурирование

ЦЕЛЕВОЙ ПОКАЗАТЕЛЬ: валовый региональный продукт в расчете на душу населения за 2015-2020 г.

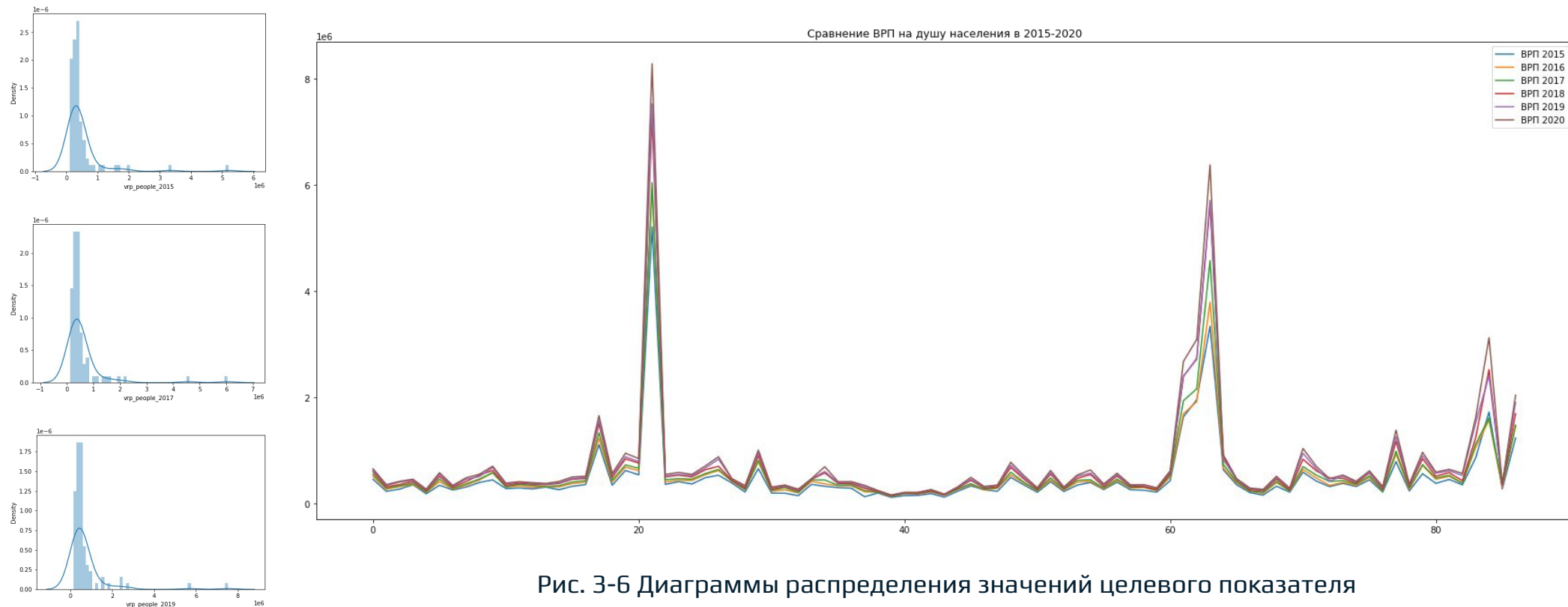


Рис. 3-6 Диаграммы распределения значений целевого показателя

ПРЕДСКАЗАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ РЕГИОНОВ РОССИИ

Этап 4: алгоритмы машинного обучения

ЗАДАНИЕ ДЛЯ ОБУЧЕНИЯ, ИСХОДНЫЕ ДАННЫЕ ДЛЯ ВЫБОРА И ПРИМЕНЕНИЯ МОДЕЛИ:

- **Алгоритм регрессии**, работающий на небольших объемах данных, настройка гиперпараметров по ходу обучения
- Обучение алгоритма на признаках за **2015-2017** годы с целевым результатом за **2019**, так как результат деятельности органов власти по социально-экономическому развитию носит отложенный по времени характер
- Алгоритм машинного обучения не учитывает геополитические и иные риски, влияющие на социально-экономическое развитие, за исключением рисков чрезвычайных ситуаций, учтенных в признаках
- В результате применения алгоритмов машинного обучения должны быть получены предсказания по целевому результату в **2022** для каждого из объектов выборки по признакам **2018-2020** годов, т.е. по результатам уже проделанной работы

ПРЕДСКАЗАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ РЕГИОНОВ РОССИИ

Этап 4: алгоритмы машинного обучения

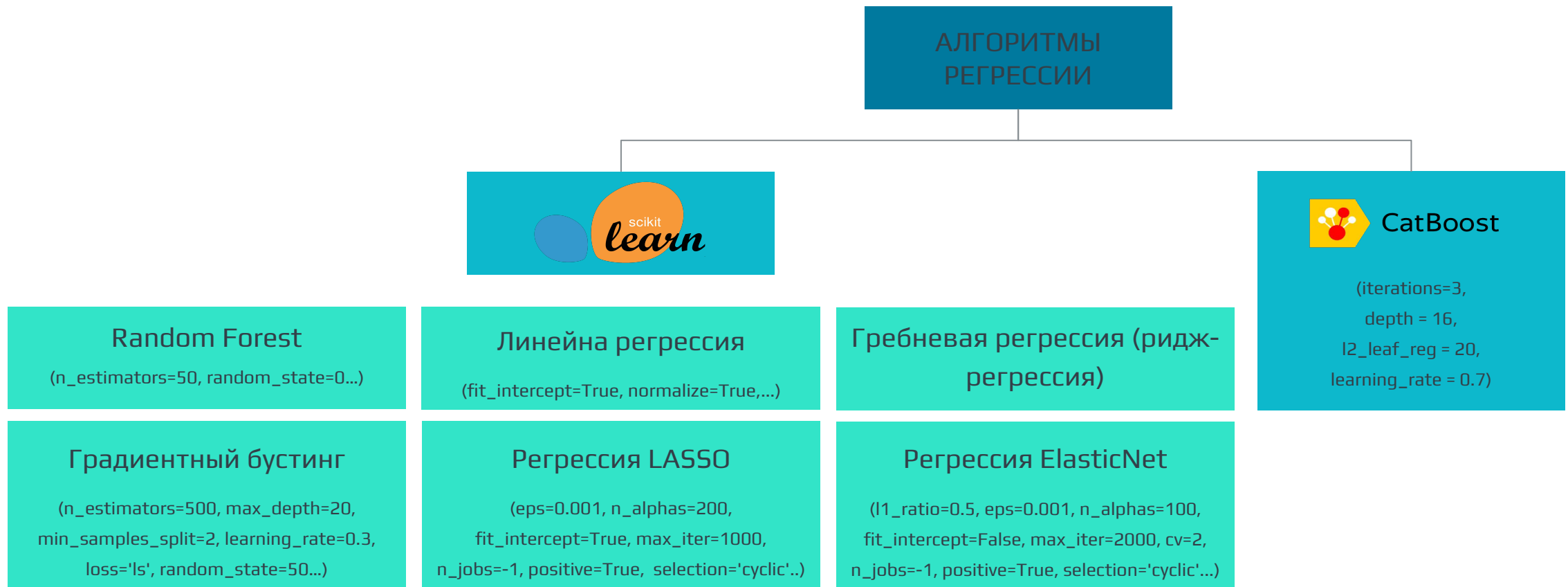


Рис. 7 Схема применяемых алгоритмов ML

ПРЕДСКАЗАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ РЕГИОНОВ РОССИИ

Этап 5: оценка работы алгоритмов

ИСПОЛЬЗУЕМЫЕ МЕТРИКИ:

1. **RMSE** - среднеквадратичная ошибка
2. **MAE** - средняя абсолютная ошибка
3. **R2** - коэффициент детерминации
4. **MAPE** - средняя абсолютная процентная ошибка

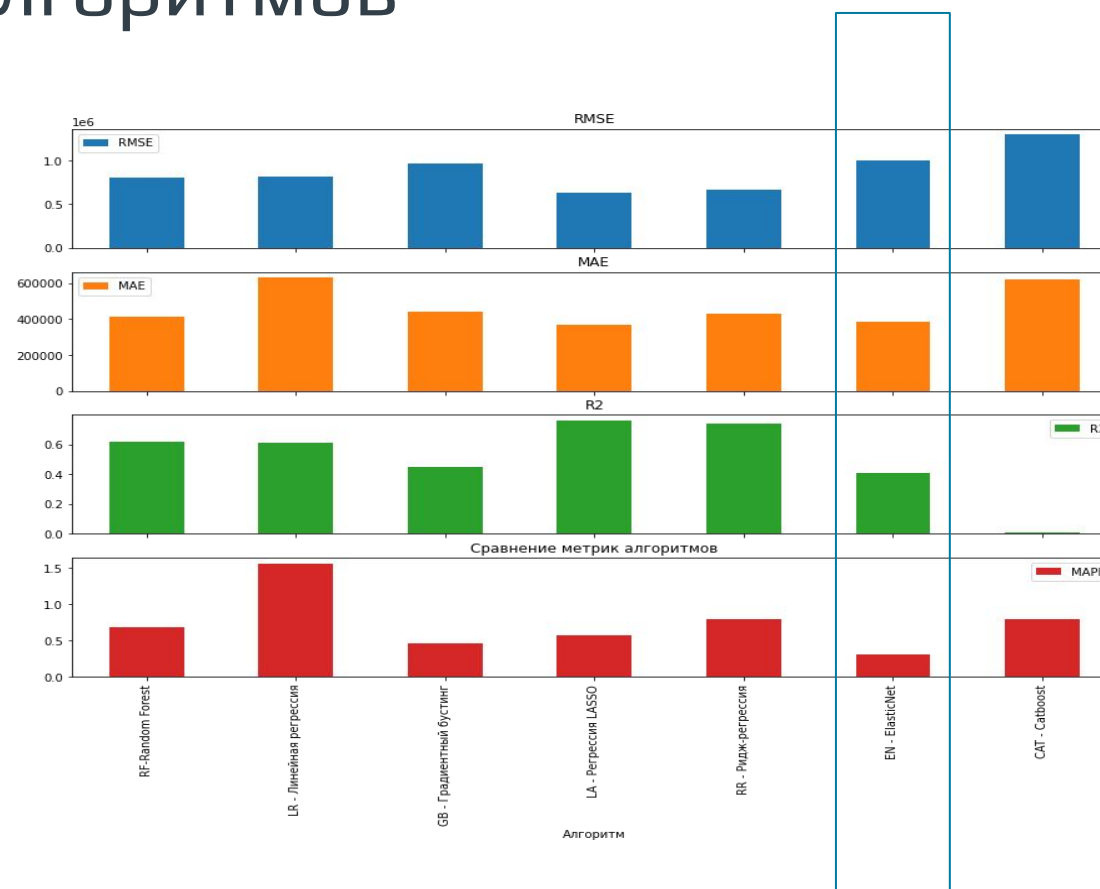


Рис. 8 Диаграмма результатов применения алгоритмов ML

ПРЕДСКАЗАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ РЕГИОНОВ РОССИИ

Этап 6: формирование результата

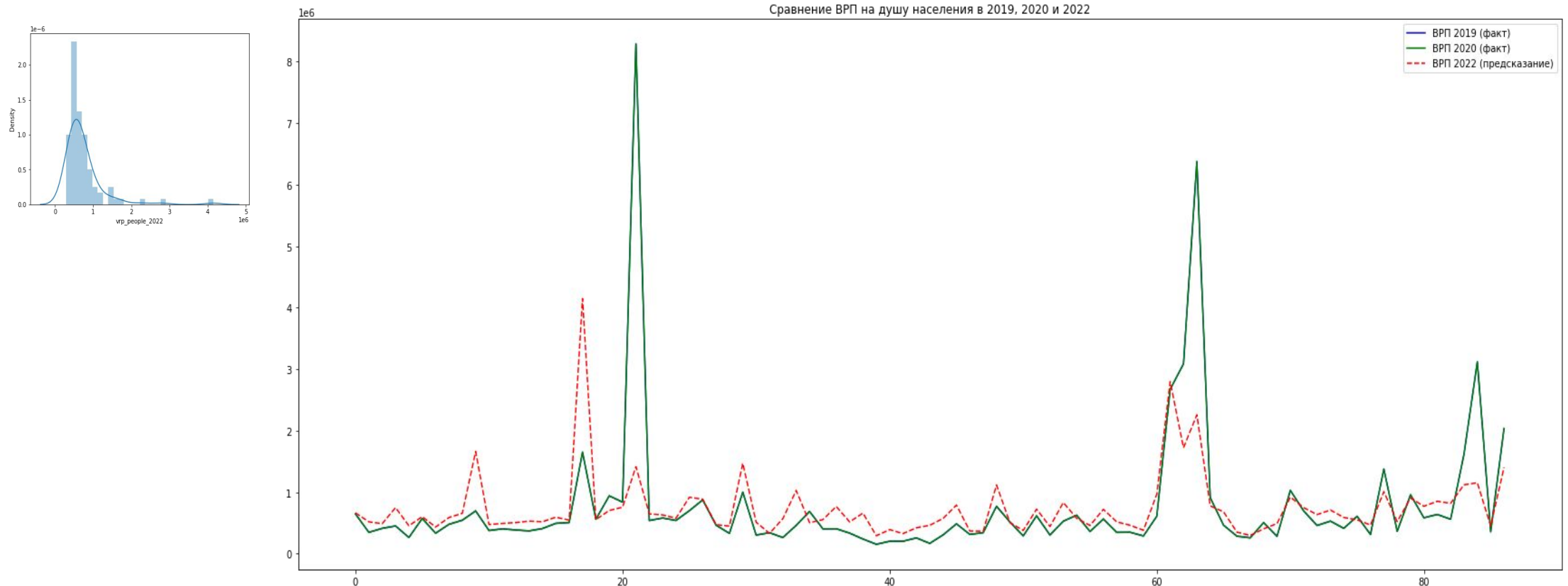


Рис. 9, 10 Диаграммы распределения значений предсказанного целевого показателя

ПРЕДСКАЗАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ РЕГИОНОВ РОССИИ

Этап 6: интерпритация результата

374 533.82 РУБ

Величина средней абсолютной ошибки, при диапазоне значений показателя предсказания от 245 411.7 до 5 710 467.4 и медианой 826 213.36 руб.

33%

Средняя абсолютная процентная ошибка предсказания на тестовой выборке

Выводы

- Поставленные цели и задачи проекта достигнуты в полном объеме.
- В результате проекта отработаны методы сбора, анализа и работы с данными, на практике применены алгоритмы машинного обучения.
- По мнению Минэкономразвития Новосибирской области, предложенная модель, может быть использована на практике в качестве альтернативной системы прогнозирования показателей социально-экономического развития.

Принципы работы модели могут быть использованы в иных сферах экономического прогнозирования, в том числе оперативного мониторинга эффективности тех или иных мероприятий. Для внедрения модели в качестве основной необходимо повышения качества предсказания, снижение ошибок работы, обучение специалистов работе с данными.

- Для снижения ошибок и повышения качества работы алгоритма целесообразно увеличение объема данных (признаков).

Список использованных источников

1. Цели в области устойчивого развития // Интернет-ссылка: <https://www.un.org/sustainabledevelopment/ru/sustainable-development-goals/>
2. О целях устойчивого развития // Интернет-ссылка: <https://rosstat.gov.ru/sdg>
3. Статистический ежегодник «Цели устойчивого развития в Российской Федерации» // Интернет ссылка <https://rosstat.gov.ru/sdg/report/document/69771>
4. Росстат. Официальная статистика // Интернет-ссылка: <https://rosstat.gov.ru/folder/10705>
5. ЕМИСС // Интернет-ссылка: <https://www.fedstat.ru/>
6. 3.3. Metrics and scoring: quantifying the quality of predictions // Интернет-ссылка: https://scikit-learn.org/stable/modules/model_evaluation.html
7. Метрики для оценки алгоритмов машинного обучения в Python // Интернет-ссылка <https://www.machinelearningmastery.ru/metrics-evaluate-machine-learning-algorithms-python/>
8. Seaborn для визуализации данных в Python // Интернет-ссылка <https://pythonru.com/biblioteki/seaborn-plot>
9. Scikit-learn // Интернет-ссылка https://runebook.dev/ru/docs/scikit_learn/

Список использованных источников

10. оттенков matplotlib — The Master Plots (с полным кодом на Python) // Интернет-ссылка
<https://habr.com/ru/post/468295/>
11. 5 алгоритмов регрессии в машинном обучении, о которых вам следует знать // Интернет-ссылка:
<https://habr.com/ru/company/vk/blog/513842/>
12. Методы регрессионного анализа в Data Science // Интернет-ссылка: <https://habr.com/ru/company/otus/blog/485972/>
13. Базовые принципы машинного обучения на примере линейной регрессии // Интернет-ссылка:
<https://habr.com/ru/company/ods/blog/322076/>
14. Тренды в области устойчивого развития // Интернет-ссылка:
https://partner.sberuniversity.online/activity/246759/view?utm_source=notifications&utm_medium=email&utm_campaign=knowledge_base_news