ПОЛИТЕХ
Санкт-Петербургский политехнический университет Петра Великого

ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА

СИНТЕЗ МОДЕЛИ ДЛЯ ПРОГНОЗА РОЖДАЕМОСТИ В РАЗНЫХ СУБЪЕКТАХ РОССИЙСКОЙ ФЕДЕРАЦИИ

Билев Севастьян Николаевич

научный руководитель
Семендяев Родион Юрьевич

19.02.2026

ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА

Актуальность темы

Демографические процессы, в частности рождаемость, являются ключевым фактором социально-экономического развития и национальной безопасности России.

Существенные региональные различия в уровнях и динамике рождаемости требуют разработки адресных прогнозных моделей для обоснования эффективной демографической политики.

ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА

Цели и задачи

Цель работы:

Разработка и реализация программной прогнозной модели уровня рождаемости (в виде общего коэффициента рождаемости или иных метрик) для субъектов Российской Федерации с использованием языка программирования Python

Задачи исследования:

- ❑ Анализ предметной области
- ❑ Сбор и предобработка данных
- ❑ Разработать и обучить модели
- ❑ Оценить точность и верифицировать результаты

Сравнить с базовыми алгоритмами и прогнозами Росстата

ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА

Объекты и предмет исследования

Объект исследования

Демографические процессы в субъектах Российской Федерации (динамика уровня рождаемости).

Предмет исследования

Математические методы и алгоритмы машинного обучения для прогнозирования на основе панельных данных.

Методы исследования

Анализ данных (EDA), статистические методы, машинное обучение

Характеристики датасета

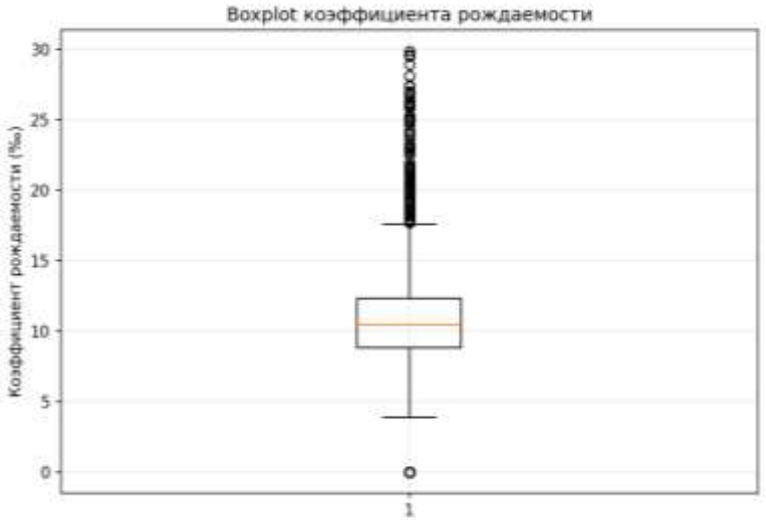
Период: 1990 – 2023 гг. (34 года)
Регионы: 85 субъектов РФ
Формат: Панельные данные (регионы × годы)
Объем: 3094 наблюдения, 39 показателей

Целевая переменная

общий коэффициент рождаемости, ‰ (crude_birth_rate_per_1000)

ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА

Разведочный анализ данных



ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА

Подготовка данных и формирование признаков

Шаг 1. Отбор признаков

Выбраны признаки с минимальным числом пропусков (total_population, share_of_urban_population, avg_per_capita_income_rub, maternity_capital_flag и др.).

Шаг 2. Создание лагов

Созданы лаги целевой переменной (t-1, t-2, t-3) для учета автокорреляции.

Шаг 3. Кодирование регионов

LabelEncoder для учета региональной специфики.

Шаг 4. Масштабирование и преобразование

StandardScaler для признаков, log(1+y) для целевой переменной (для стабилизации дисперсии).

Шаг 5. Формирование выборок

Хронологическое разделение (обучение: 1990-2017, тест: 2018-2023). Создание последовательностей (окно = 3 года) для LSTM.

ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА

Архитектура LSTM-модели

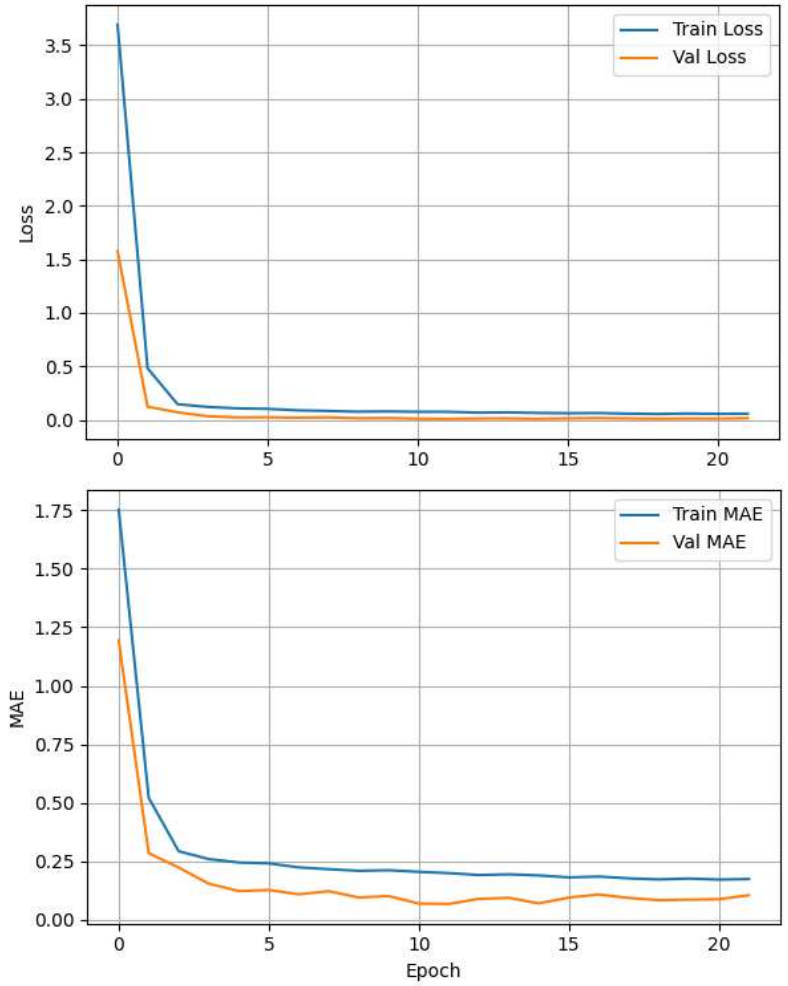
Слой (тип)	Выходная размерность	Количество параметров
LSTM (1)	(None, 3, 64)	18 688
Dropout	(None, 3, 64)	0
LSTM (2)	(None, 32)	12 416
Dropout	(None, 32)	0
Dense	(None, 16)	528
Dense	(None, 1)	17
Всего		31 649

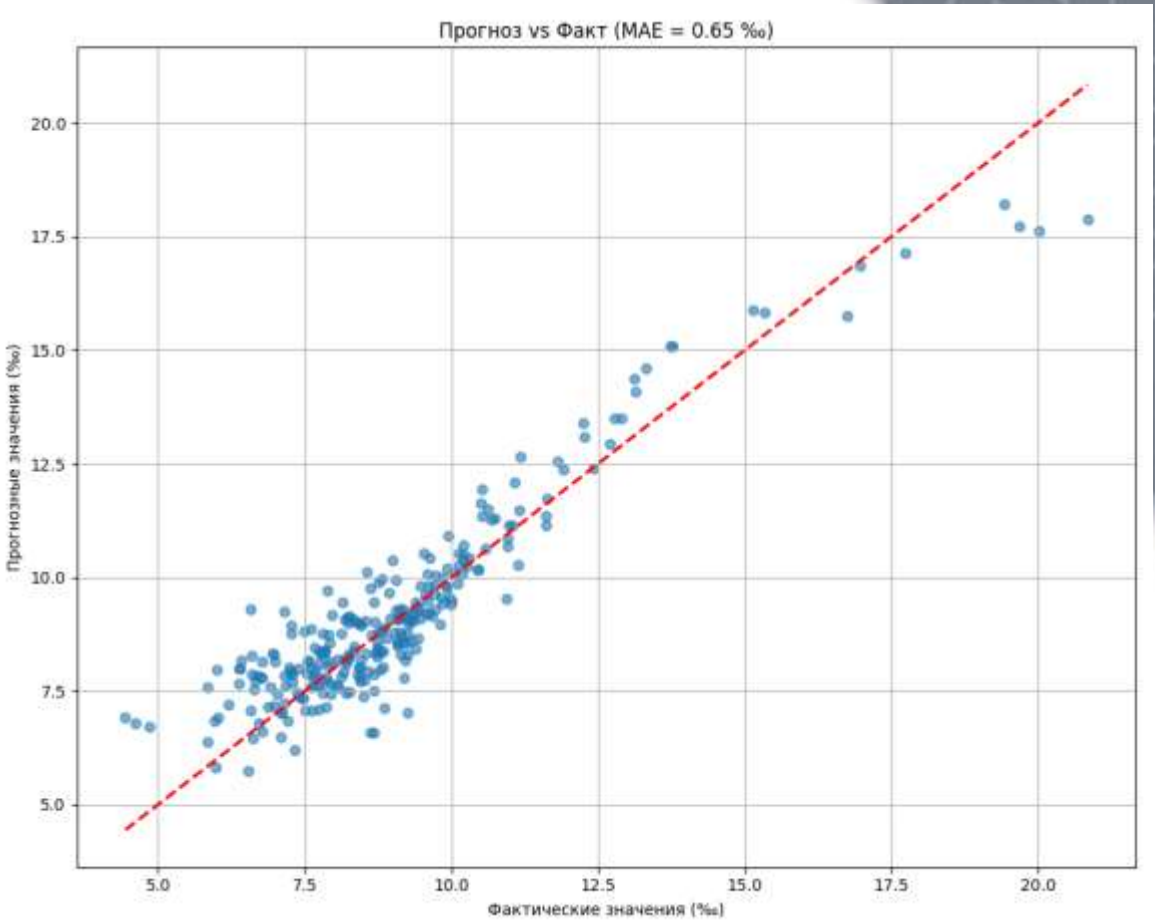
ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА

Результаты обучения и оценка модели

Метрики на тестовой выборке

- ❑ MAE = 0.609 ‰
- ❑ RMSE = 0.774 ‰
- ❑ R^2 = 0.890





ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА

Сравнение с альтернативными подходами

Сравнение MAE (‰)

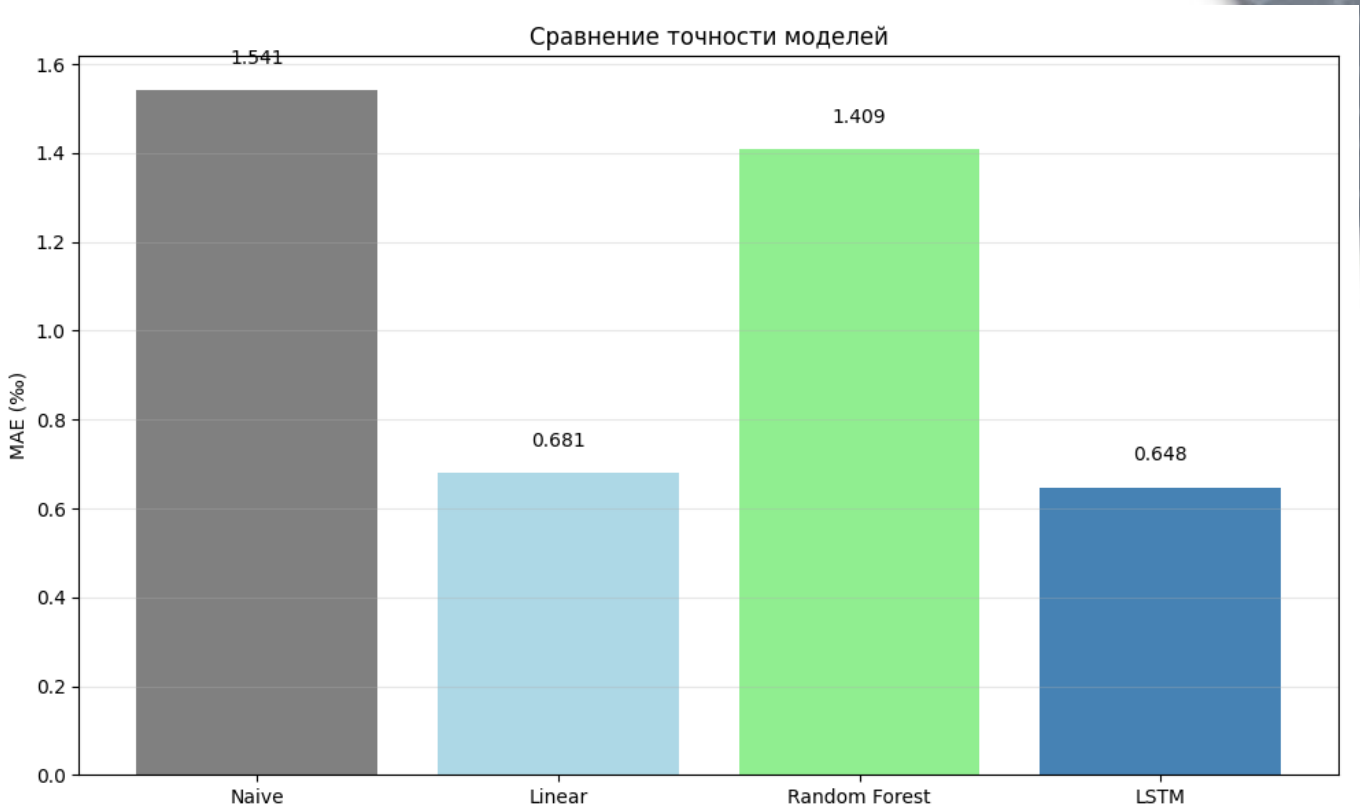
Наивный прогноз – 1.541 ‰

Случайный лес – 1.409 ‰

Линейная регрессия – 0.681 ‰

LSTM – 0.609 ‰

LSTM превосходит линейную регрессию на **10.7%**, случайный лес на **56.8%**, наивный прогноз на **60.5%**



ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА

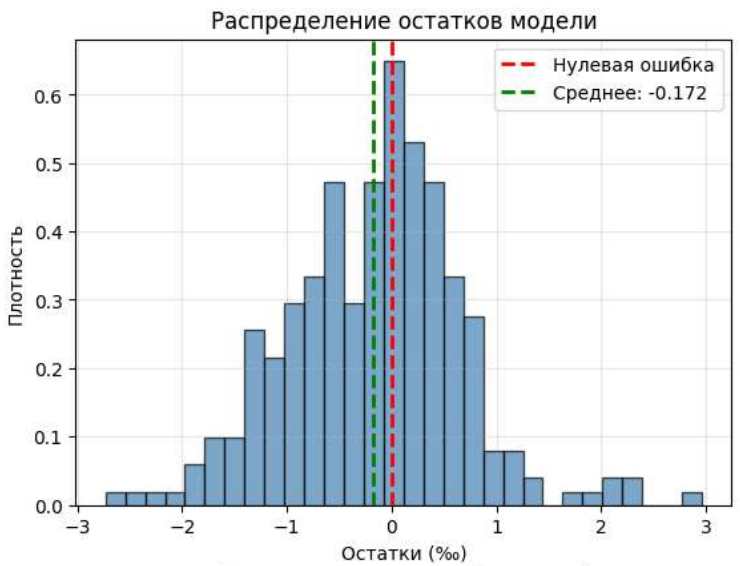
Анализ ошибок и верификация модели

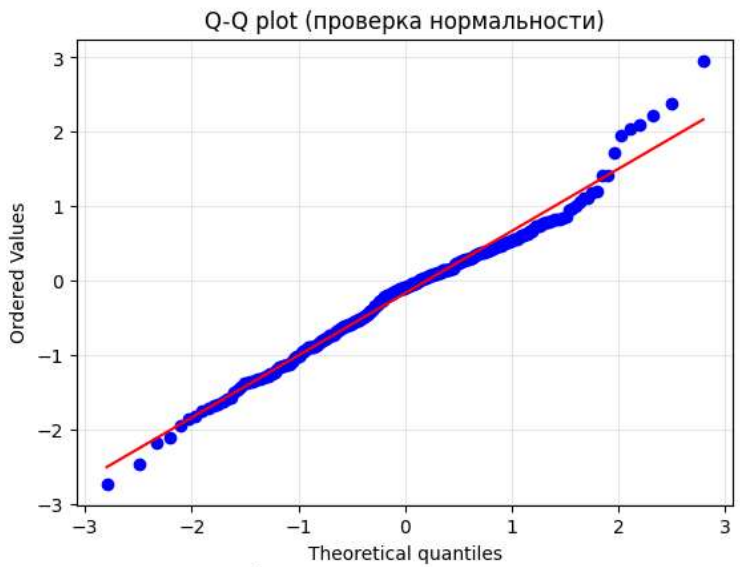
MAE < 0.5‰
Центральная Россия, Северо-Запад, Поволжье

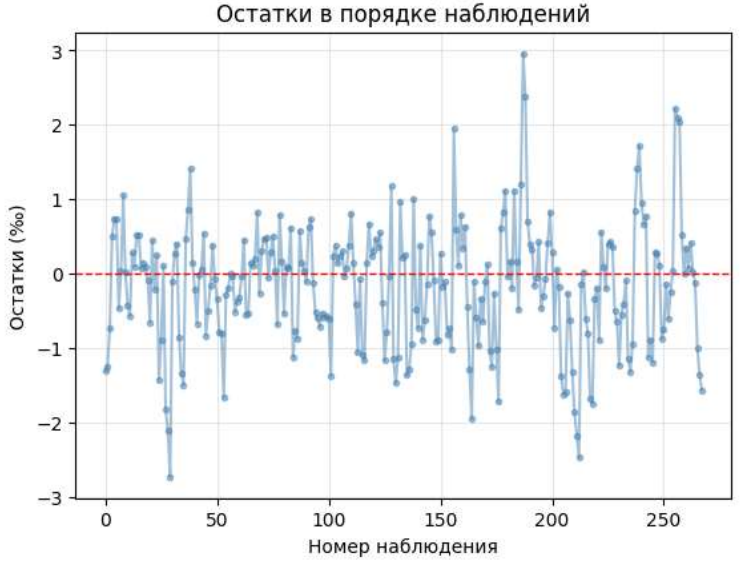
0.5‰ ≤ MAE < 1.0‰
Большинство регионов

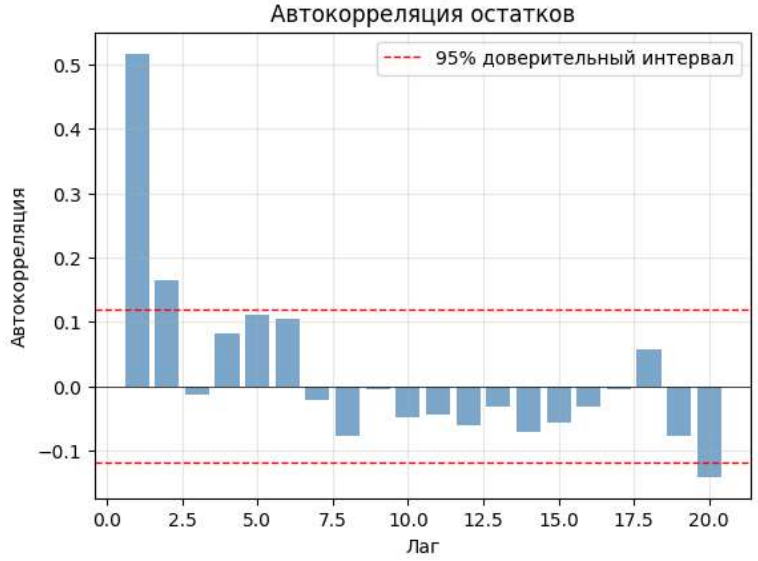
MAE ≥ 1.0‰
Северный Кавказ, Тыва

Модель адекватна
остатки распределены нормально, нет систематической ошибки, но имеет ограничения для регионов с высокой волатильностью









ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА

Анализ ошибок и верификация модели

Год	Факт (‰)	Прогноз Росстата (‰)	Прогноз LSTM (‰)	Ошибка Росстата (‰)	Ошибка LSTM (‰)
2020	9,73	10,10	9,91	+0,37	+0,18
2021	9,50	9,85	9,68	+0,35	+0,18
2022	8,89	9,30	9,12	+0,41	+0,23
2023	8,65	9,05	8,89	+0,40	+0,24

Разработанная LSTM-модель демонстрирует более высокую точность прогноза по сравнению с официальными прогнозами Росстата: средняя ошибка LSTM составляет 0,21 ‰ против 0,38 ‰ у Росстата

ВЫСШАЯ ИНЖЕНЕРНАЯ ШКОЛА

Заключение и итоги работы

Основные итоги

- ❑ Разработана и реализована прогнозная LSTM-модель на Python
- ❑ Проведен глубокий анализ региональных демографических данных
- ❑ Достигнута высокая точность прогноза (MAE=0.609‰, R^2=0.890)
- ❑ Модель превосходит базовые алгоритмы и официальные прогнозы Росстата

СПАСИБО
ЗА ВНИМАНИЕ!

ПОЛИТЕХ
Санкт-Петербургский
политехнический университет
Петра Великого

ВЫСШАЯ
ИНЖЕНЕРНАЯ
ШКОЛА

19.02.2026