

Научная статья
УДК 004
<https://doi.org/10.24143/2072-9502-2026-2-61-73>
EDN IFWTJP

Модульный подход к разработке информационной системы для прогноза цен на авиабилеты

Татьяна Евгеньевна Саратова¹,
Илья Сергеевич Кожевников^{2✉}, Владислав Андреевич Шульгин³

^{1, 3}*МИРЭА – Российский технологический университет,*
Москва, Россия, is.kozhevnikov@gmail.com✉

²*ООО «Айти класс»,*
Москва, Россия

Аннотация. Одним из вызовов для компаний на рынке авиаперевозок является изменчивость цен на авиабилеты, что затрудняет планирование и принятие решений. Исследованы подходы и технологии прогнозирования цен на авиабилеты с учетом различных параметров. Описывается интегрированная информационная система, предназначенная для краткосрочного прогнозирования цен на авиабилеты по популярным направлениям. Модуль сбора данных агрегирует продажи, показатели спроса, сезонные факторы и операционные параметры авиакомпаний, сохраняя их в базе PostgreSQL. На аналитическом уровне рассматриваются четыре подхода: линейная регрессия, градиентный бустинг, Prophet и рекуррентная сеть LSTM. Проведен сравнительный эксперимент на выборке из пяти крупнейших маршрутов: у LSTM средние ошибки составили RMSE 0,065–0,153 и MAE 0,049–0,117, что оказалось более высоким результатом при сохранении устойчивости к выбросам. Архитектурно система разделена на модуль машинного обучения (Python + VS Code), хранилище данных (PostgreSQL) и слой визуализации (Power BI); диаграмма развертывания демонстрирует взаимодействие компонентов через REST-API. В пользовательском интерфейсе доступны график цен и календарь минимальных тарифов на ближайшие пять дней. Представленные результаты подтверждают применимость LSTM для динамического ценообразования и показывают возможности масштабирования методологии на финансы, энергетику и логистику, где требуются прогнозы временных рядов с высоким показателем волатильности. Обозначены направления дальнейших исследований и пути совершенствования технологий прогнозирования для повышения точности и устойчивости решений в отрасли.

Ключевые слова: прогнозирование, машинное обучение, нейросети, архитектура системы, LSTM, авиабилеты

Для цитирования: *Саратова Т. Е., Кожевников И. С., Шульгин В. А.* Модульный подход к разработке информационной системы для прогноза цен на авиабилеты // *Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика.* 2026. № 2. С. 61–73. <https://doi.org/10.24143/2072-9502-2026-2-61-73>. EDN IFWTJP.

Original article

Modular approach to the development of an information system for forecasting airline ticket prices

Tatyana E. Saratova¹, Ilya S. Kozhevnikov^{2✉}, Vladislav A. Shulgin³

^{1, 3}*MIREA – Russian Technological University,*
Moscow, Russia, is.kozhevnikov@gmail.com✉

²*“IT-Class” LLC,*
Moscow, Russia

Abstract. In the dynamic airline market, one of the main challenges for companies is the high volatility of ticket prices, which complicates planning and decision-making. This paper investigates approaches and technologies for forecasting airline ticket prices under multiple parameters. It describes an integrated information system designed for short-term price prediction on popular routes. The data-collection module aggregates historical sales, demand indicators,

seasonal factors, and airline operating parameters, storing them in a PostgreSQL database. Four forecasting approaches are examined at the analytics layer-linear regression, gradient boosting, Prophet, and a Long Short-Term Memory (LSTM) recurrent network. A comparative experiment on five major routes shows that LSTM achieves the best performance, with average errors of RMSE 0.065-0.153 and MAE 0.049-0.117, while remaining robust to outliers. Architecturally, the system is divided into a machine-learning module (Python + VS Code), a data repository (PostgreSQL), and a visualization layer (Power BI); the deployment diagram illustrates component interaction via a REST API. The user interface offers an interactive price chart and a calendar of minimum fares for the next five days. The reported results confirm the suitability of LSTM for dynamic pricing and demonstrate the scalability of the methodology to finance, energy, and logistics domains where highly-volatile time-series forecasting is required. The conclusion outlines directions for further research and ways to improve forecasting technologies in order to enhance the accuracy and resilience of decision-making in the industry.

Keywords: forecasting, machine learning, neural networks, system architecture, LSTM, air tickets

For citation: Saratova T. E., Kozhevnikov I. S., Shulgin V. A. Modular approach to the development of an information system for forecasting airline ticket prices. *Vestnik of Astrakhan State Technical University. Series: Management, computer science and informatics*. 2026;2:61-73. (In Russ.). <https://doi.org/10.24143/2072-9502-2026-2-61-73>. EDN IFWTJP.

Введение

Рынок авиаперевозок является непрерывно развивающейся отраслью экономики и характеризуется высокой динамичностью и интенсивной конкуренцией, что делает процесс ценообразования авиабилетов сложным и волатильным. Объем мирового рынка авиаперевозок оценивается в 110,91 млрд долл. США в 2024 г. и, по прогнозам, способен достичь 147,35 млрд долл. США к 2029 г., при этом среднегодовой темп роста составит 4 % в течение прогнозируемого периода (2024–2029 гг.). Данные показатели подчеркивают высокую финансовую емкость рынка, где даже незначительное повышение точности ценообразования с помощью машинного обучения способно принести компаниям существенный экономический эффект.

Цены на авиабилеты подвержены колебаниям, зависящим от множества факторов, таких как сезонность, наличие специальных предложений и изменения цен со стороны авиакомпаний. В условиях быстро меняющейся ситуации важно не только оперативно формировать предложение, но и уметь предсказывать, как изменится цена на него в будущем [1–4]. Корректные технологии и подходы к прогнозированию цен на авиабилеты по заданным параметрам могут стать решением в вопросах оптимизации процесса ценообразования и повышения уровня удовлетворенности клиентов. Архитектура системы, в основе которой лежат данные методы, сможет предоставить пользователям необходимую информацию о том, когда выгодно предлагать билеты клиентам, а также проанализировать ряд параметров, влияющих на их стоимость. Прогнозирование цен на авиабилеты предоставляет компаниям новые возможности для адаптации тарифной политики к колебаниям спроса, что позволяет получать дополнительную прибыль. В данной работе рассматриваются архитектура информационной системы прогнозирования, принципы ее функционирования, а также перспективы ее коммерческого внед-

рения на рынке. Использование разработанной системы на базе машинного обучения позволяет повысить гибкость динамического ценообразования.

Целью работы является разработка информационной системы для краткосрочного прогнозирования цен на авиабилеты с низким показателем ошибки (MAPE ниже 5 %) с модульной архитектурой и возможностью масштабирования на другие отрасли.

Обзор существующих решений

Главная проблема, с которой сталкиваются пассажиры – колебания стоимости билетов в зависимости от множества факторов. Без информационной системы пользователь не может определить, когда лучше совершить покупку, рискуя либо переплатить, либо упустить выгодное предложение. Для пассажиров сервис прогнозирования цен позволяет:

- рассчитывать прогноз колебания цен на год, а также выгодно планировать поездки;
- сократить затраты на приобретение билетов;
- снижать неопределенность и стресс, связанные с выбором времени покупки.

Авиакомпания, в свою очередь, получает ряд преимуществ:

- в зависимости от прогнозов может формировать акции;
- в зависимости от прогноза формировать бюджетный план на будущий год;
- оптимизировать загрузку рейсов, снижать количество пустых мест;
- повышать прозрачность предложений для пользователей, что снижает неопределенность при покупке авиабилетов и способствует росту лояльности клиентов к бренду.

Рассмотрим практики прогнозирования цен на примере трех систем для пассажиров:

- Норрег – система, использующая машинное обучение, включая нейросети, для предсказания динамики цен и рекомендаций по покупке билетов. В качестве метода используется анализ данных,

сезонности, спроса. Ограничения системы заключаются в закрытости алгоритмов и отсутствии интеграции с данными авиакомпаний во времени;

– Google Flights – система, функционал которой включает графики цен, предупреждения о снижении стоимости. Среди методов можно выделить статистический анализ трендов. Ограничением является отсутствие долгосрочного (более 3 мес) прогнозирования;

– Kayak – система, строящая прогноз «стоит ли ждать снижения цены» с использованием регрессионных моделей. Ограничением является точность для краткосрочных прогнозов.

Среди корпоративных систем для авиакомпаний можно выделить следующие:

– PROS Revenue Management – корпоративная система, использующая динамическое ценообразование на основе спроса. К ее методам относятся градиентный бустинг, а также оптимизация. Ограничения заключаются в стоимости, требовании сложной

интеграции;

– Sabre AirPrice использует анализ конкурентных цен и прогнозирование. Методы включают правила на основе бизнес-логики. Ограничения – слабая адаптация к нелинейным трендам.

Основное решение проблемы нестабильности цен заключается в разработке информационной системы, способной прогнозировать стоимость авиабилетов на основе различных факторов: сезонности, спроса, динамики цен в прошлом, политики авиакомпаний. Для этого необходимо провести сравнительный анализ методов прогнозирования и оценки моделей.

Обзор экспериментальных данных

Экспериментальные данные, с использованием которых проводилась оценка применения моделей LSTM, Prophet, XGBoost, линейной регрессии, представлена в табл. 1.

Таблица 1

Table 1

Экспериментальные данные

Experimental data

| Ключевые поля | Расшифровка | Тип данных | Описание |
|------------------|--------------------------|-----------------|---|
| amount_total_sum | Общая сумма | DECIMAL/NUMERIC | Полная стоимость билета (включая таксы и сборы) |
| fare_sum | Стоимость тарифа | | Базовая цена билета без дополнительных сборов |
| service_fee_sum | Сбор за обслуживание | | Дополнительные сервисные сборы (может быть отрицательным при скидках) |
| tax_sum | Налоги | | Сумма налогов, включенных в стоимость |
| cost_price | Себестоимость | | Расходы авиакомпании на рейс |
| created | Дата создания записи | TIMESTAMP | Дата и время бронирования |
| status_id | Идентификатор статуса | INTEGER | Код статуса брони (например, 2 – подтверждено) |
| city_dep_name | Город вылета | VARCHAR/TEXT | Название города отправления |
| city_arr_name | Город прилета | | Название города назначения |
| aircraft | Тип самолета | | Модель воздушного судна (например, Airbus A321). |
| direction | Направление | BOOLEAN/INTEGER | 0 – прямой рейс, 1 – обратный |
| flight_duration | Длительность полета, мин | INTEGER | Время в пути |
| carrier_code | Код авиакомпании | VARCHAR(3) | IATA-код перевозчика (например, TK – Turkish Airlines) |
| baggage | Норма багажа | VARCHAR/TEXT | Условия провоза багажа (например, 1P23K – 1 место до 23 кг) |
| cbaggage | Норма ручной клади | | Условия провоза ручной клади (например, 1P8K – 1 место до 8 кг) |
| change | Возможность изменения | BOOLEAN/INTEGER | 1 – разрешено, 0 – запрещено |
| refund | Возможность возврата | | |

На рис. 1 представлен фрагмент исторических данных о ценах на авиабилеты, объем выборки сос-

тавляет 1 004 844 записей.

| amount_total_sum | fare_sum | service_fee_sum | tax_sum | cost_price | created | status_id | dep_nam | arr_nam | aircraft | direction | ght_durati | carrier_cod | baggage | cbaggage | change | refund |
|------------------|----------|-----------------|----------|------------|---------------------|-----------|-----------------|-----------------|------------|-----------|------------|-------------|---------|----------|--------|--------|
| 85515 | 63030,6 | 4888,52 | 17595,88 | 82063,68 | 2026-05-12 23:58:49 | 2 | Белград | Стамбул | Airbus A32 | 0 | 115 TK | 1P23K | 1P8K | | 1 | 0 |
| 85515 | 63030,6 | 4888,52 | 17595,88 | 82063,68 | 2026-05-12 23:58:49 | 2 | Москва | Стамбул | Airbus A32 | 1 | 310 S7 | 1PC | 1P10K | | 1 | 1 |
| 85515 | 63030,6 | 4888,52 | 17595,88 | 82063,68 | 2026-05-12 23:58:49 | 2 | Стамбул | Белград | Boeing 737 | 1 | 105 TK | 1P23K | 1P8K | | 1 | 0 |
| 85515 | 63030,6 | 4888,52 | 17595,88 | 82063,68 | 2026-05-12 23:58:49 | 2 | Стамбул | Москва | Boeing 737 | 0 | 280 S7 | 1PC | 1P10K | | 1 | 1 |
| 29331 | 25740 | -369 | 3960 | 28985,14 | 2026-05-12 23:49:31 | 2 | Москва | Ош | Airbus A32 | 0 | 265 KA | 1P20K | 1P5K | | 1 | 1 |
| 6382 | 5330 | 545 | 507 | 6102,87 | 2026-05-12 23:47:26 | 2 | Санкт-Петербург | Москва | Airbus A32 | 0 | 90 SU | 0PC | 1P10K | | 1 | 0 |
| 128575 | 95440 | -2273 | 35408 | 128380,8 | 2026-05-12 23:41:03 | 2 | Париж | Стамбул | Airbus A32 | 1 | 220 TK | 1P30K | 1P8K | | 1 | 0 |
| 128575 | 95440 | -2273 | 35408 | 128380,8 | 2026-05-12 23:41:03 | 2 | Стамбул | Санкт-Петербург | Airbus A32 | 1 | 220 TK | 1P30K | 1P8K | | 1 | 0 |
| 128575 | 95440 | -2273 | 35408 | 128380,8 | 2026-05-12 23:41:03 | 2 | Санкт-Петербург | Стамбул | Airbus A32 | 0 | 240 TK | 1P30K | 1P8K | | 1 | 1 |
| 128575 | 95440 | -2273 | 35408 | 128380,8 | 2026-05-12 23:41:03 | 2 | Стамбул | Париж | Airbus A32 | 0 | 220 TK | 1P30K | 1P8K | | 1 | 1 |
| 16833 | 15273,65 | 1559,35 | 0 | 16205,47 | 2026-05-12 23:38:33 | 2 | Стамбул | Джидда | | 0 | 240 FZ | 0PC | 1P10K | | 0 | 0 |
| 29640 | 13565 | 522 | 15553 | 29284,87 | 2026-05-12 23:37:11 | 2 | Москва | Ташкент | Boeing 787 | 0 | 225 HY | 1PC | 1P8K | | 1 | 1 |

Рис. 1. Фрагмент исторических данных о ценах на авиабилеты

Fig. 1. A fragment of historical data on airline ticket prices

Для сравнения моделей использовались исторические данные о ценах на авиабилеты по пяти популярным маршрутам: Москва – Ереван, Москва – Бишкек, Москва – Стамбул, Москва – Ош, Санкт-Петербург – Москва. Период анализа составил 2023–2024 гг. Данные были разделены на обучающую выборку: 70 % (январь 2023 г. – сентябрь 2024 г.) и тестовую выборку: 30 % (октябрь–декабрь 2024 г.).

Анализ методов прогнозирования и оценки моделей

Для прогнозирования динамики цен рассматриваются следующие математические подходы.

Линейная регрессия используется для выявления общей тенденции к изменению цен на основе временного ряда. Расчет производится по формуле [5]

$$P(t) = \beta_0 + \beta_1 t + \epsilon, \quad (1)$$

где $P(t)$ – прогнозируемая цена в момент времени t ; β_0 , β_1 – параметры модели; ϵ – ошибка предсказания.

Градиентный бустинг (XGBoost, LightGBM) строит ансамбль деревьев решений, что позволяет учитывать нелинейные зависимости. Ниже приведена функция прогноза для градиентного бустинга:

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x), \quad (2)$$

где $F_m(x)$ – прогноз на m -й итерации; $h_m(x)$ – новое дерево решений; γ – коэффициент обучения.

LSTM – это разновидность рекуррентных нейронных сетей, хорошо подходящая для анализа временных рядов [6–8]. Обновление состояний ячейки LSTM описывается следующими формулами:

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t; \quad (3)$$

$$h_t = o_t \tan h(C_t), \quad (4)$$

где f_t , i_t , o_t – весовые коэффициенты; C_t – состояние ячейки; \tilde{C}_t – вектор-кандидат новых значений состояния ячейки; h_t – выход сети.

Prophet – статистический метод, предназначенный для моделирования сезонных данных. Базовая модель Prophet описывается формулой

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t, \quad (5)$$

где $g(t)$ – тренд; $s(t)$ – сезонность; $h(t)$ – праздничные эффекты.

На основе анализа различных методов прогнозирования цен на авиабилеты, рассмотренных ранее, выберем наиболее подходящий метод с учетом требований к информационной системе: высокая точность предсказания, способность обрабатывать сложные нелинейные зависимости, работа с временными рядами, устойчивость к выбросам, скорость обучения и предсказания. Приведем формулы и методы оценки работы информационной системы.

Средняя абсолютная процентная ошибка MAPE представлена в формуле

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100 \%, \quad (6)$$

где y_i – фактическое значение цены билета на i -м шаге; \hat{y}_i – предсказанное значение; n – количество наблюдений.

Используем коэффициент детерминации R^2 , который показывает, насколько хорошо модель объясняет вариацию цен:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}, \quad (7)$$

где \bar{y} – среднее значение фактических данных.

Линейная регрессия плохо справляется с нелинейностями ($R^2 < 0,5$).

XGBoost и Prophet показывают хорошие результаты ($R^2 \approx 0,7-0,9$).

LSTM лучше всего обрабатывает нелинейности, т. к. использует нейросетевую архитектуру ($R^2 > 0,9$).

Используем автокорреляцию остатков, представленную в формуле

$$\gamma_k = \frac{\sum (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum (y_t - \bar{y})^2}, \quad (8)$$

где γ_k – коэффициент автокорреляции с лагом k .

LSTM и Prophet лучше всего учитывают временные зависимости ($\gamma_k \approx 0,8-0,95$).

XGBoost ограничен в работе с временными рядами. Линейная регрессия игнорирует автокорреляцию (8). Также все модели могут быть оценены по MAPE (6) и R^2 (7).

Оценка скорости работы:

- линейная регрессия (1) выполняется за доли секунды, т. к. использует простые матричные операции;
- XGBoost (2) требует 5–10 с из-за построения ансамбля деревьев решений;

– LSTM (3), (4) медленнее (30–60 с на обучение) из-за сложных вычислений и рекуррентной архитектуры;

– Prophet (5) работает быстрее, чем LSTM, т. к. использует аддитивные и мультипликативные временные компоненты.

Усредненные метрики для всех маршрутов представлены в табл. 2.

Таблица 2

Table 2

Сравнительный анализ моделей прогнозирования
Comparative analysis of forecasting models

| Модель | MAPE, % | R^2 |
|--------------------|---------|-------|
| LSTM | 4,2 | 0,94 |
| Prophet | 6,8 | 0,82 |
| XGBoost | 5,9 | 0,87 |
| Линейная регрессия | 12,4 | 0,48 |

LSTM является лучшим выбором, т. к. обеспечивает высокую точность предсказаний, хорошо обрабатывает нелинейности и учитывает временные зависимости. Однако он требует больше времени на обучение по сравнению с XGBoost и Prophet.

Архитектура информационной системы

Архитектура информационной системы состоит из нескольких ключевых компонентов, которые обеспечивают корректную работу системы прогнозирования цен на авиабилеты. Основные компоненты системы включают:

1. Модуль сбора и обработки данных, который отвечает за сбор информации о ценах, сезонности, спросе и других факторах, влияющих на стоимость билетов (Visual Studio Code).

2. Модуль аналитики и машинного обучения анализирует данные и прогнозирует динамику цен,

используя выбранные алгоритмы (LSTM).

3. Модуль хранения данных накапливает основную информацию об исторических данных и прогнозах (PostgreSQL).

4. Модуль визуализации позволяет пользователям системы легко понимать прогнозируемые цены с помощью графиков и диаграмм (Power BI).

Для информационной системы, которая включает в себя модуль аналитики и машинного обучения, модуль хранения данных на PostgreSQL, модуль визуализации на Power BI, а также использует Visual Studio Code для аналитики, можно составить диаграмму развертывания, которая покажет распределение компонентов системы по серверам и взаимодействие между ними. На рис. 2 изображена схема архитектуры разрабатываемой информационной системы, представленной в диаграмме развертывания.

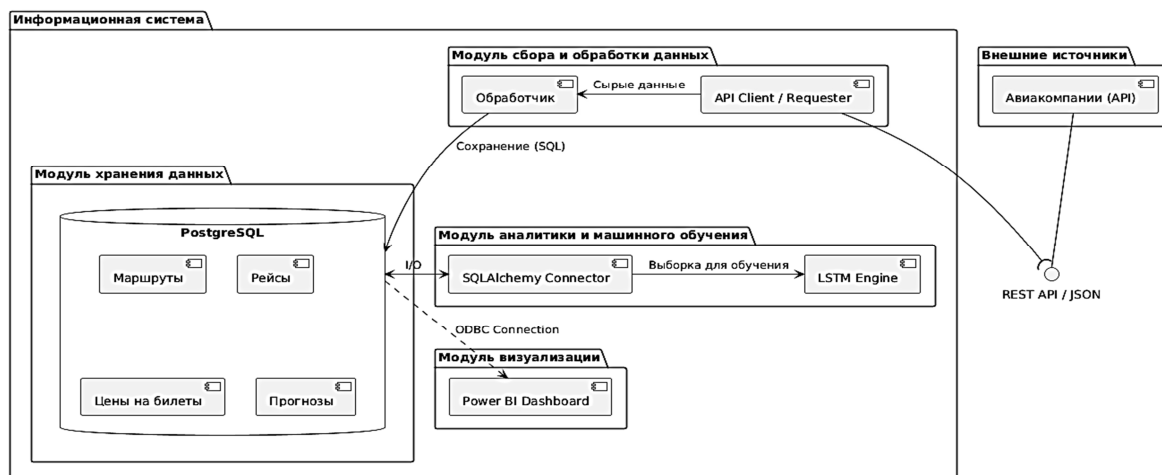


Рис. 2. Схема разрабатываемой информационной системы

Fig. 2. The scheme of the information system being developed

Реализация схемы разрабатываемой информационной системы:

1. Данные внешних API авиакомпаний предоставляют «сырую» информацию об авиабилетах в формате JSON и являются отправной точкой для всего потока данных в системе.

2. Модуль сбора и обработки данных отвечает за первичный прием и подготовку информации:

– API Client / Requester выступает в роли клиента, который отправляет запросы к внешним источникам и получает ответы;

– обработчик выполняет десериализацию полученных JSON-данных, их очистку от ошибок и нормализацию для приведения к единому формату. После обработки данные сохраняются в базу данных с помощью SQL-запросов.

3. Модуль хранения данных (PostgreSQL) – центральное хранилище системы, реализованное на базе системы управления базами данных (СУБД) PostgreSQL. Оно структурировано по ключевым сущностям:

– маршруты и рейсы: информация о географии полетов и характеристиках бортов;

– цены на билеты: массив исторических данных о стоимости (тарифы, таксы, сборы);

– прогнозы: результаты работы нейросети, сохраненные для последующей визуализации.

4. Модуль аналитики и машинного обучения (LSTM Engine) – ядро системы, разработанное на языке Python в среде Visual Studio Code. Данный модуль включает в себя два базовых компонента:

– SQLAlchemy Connector: обеспечивает интерфейс взаимодействия между кодом Python и базой данных для извлечения обучающих выборок и записи результатов;

– LSTM Engine: реализует рекуррентную нейронную сеть Long Short-Term Memory, которая анализирует временные ряды и формирует прогноз цен на ближайшие 5 дней.

5. Модуль визуализации (Power BI Dashboard) – слой представления данных для конечного пользователя. С помощью ODBC-соединения Power BI подключается напрямую к PostgreSQL, извлекает актуальные прогнозы и формирует интерактивные отчеты, графики цен и календари минимальных тарифов.

Ниже приведено описание трансформации данных на каждом этапе работы информационной системы (табл. 3).

Таблица 3

Table 3

Таблица спецификации входных и выходных параметров модулей информационной системы

Table of specifications of input and output parameters of information system modules

| Этап | Вход | Выход |
|---------------------------------------|---------------------------|---------------------------------------|
| Модуль сбора и обработки данных | Сырой JSON из внешних API | Структурированные записи в PostgreSQL |
| Модуль аналитики и машинного обучения | Исторические данные | Массив прогнозных цен на 5 дней |
| Модуль визуализации | Данные из БД (SQL) | Дашборды и графики для пользователя |

Описание технического решения

Информационная система прогнозирования цен на авиабилеты оперирует различными информационными потоками, обеспечивающими сбор, обработку, анализ и предоставление прогнозируемых данных пользователям. Данная архитектура позволяет гибко масштабировать информационную систему, интегрировать новые источники данных и обновлять алгоритмы прогнозирования [9]. На рынке представлены следующие решения:

1. Visual Studio Code (VS Code) – бесплатный кроссплатформенный редактор кода с поддержкой множества языков программирования и расширений. Обладает удобным интерфейсом, встроенным терминалом и системой отладки. Особенности VS Code:

– поддержка Python, JavaScript, PHP, C++, C#, HTML, CSS и др.;

– совместимость с Windows, macOS и Linux.

2. JetBrains PyCharm (PyCharm) – мощная среда разработки, ориентированная на работу с Python. Обладает встроенной поддержкой анализа кода, тестирования, отладки и управления зависимостями. Особенности:

– поддержка Python, HTML, CSS, JavaScript;

– совместимость с Windows, macOS и Linux.

3. Google Colab – инструмент для интерактивного программирования, часто используемый для анализа данных и машинного обучения. Позволяет создавать исполняемые блокноты с кодом, графиками и пояснениями. Особенности:

– поддержка Python;

– совместимость с Windows, macOS и Linux.

Сравнительный анализ средств разработки представлен в табл. 4.

Таблица 4

Table 4

Сравнительный анализ средств разработки

Comparative analysis of development tools

| Средство разработки | Стоимость | Доступные платформы | Поддерживаемые языки |
|---------------------|-------------|-----------------------|--|
| Visual Studio Code | Бесплатно | Windows, macOS, Linux | Python, PHP, JavaScript, C++, C# и др. |
| PyCharm | От 89 долл. | | Python, HTML, CSS, JavaScript |
| Google Colab | Бесплатно | | Python |

В результате сравнительного анализа средств разработки выбран бесплатный кроссплатформенный редактор кода Visual Studio Code, т. к. он адаптирован к широкому спектру поддерживаемых языков и сов-

местим с множеством операционных систем.

Описание и сравнение систем управления базами данных (СУБД), подходящих для реализации модуля хранения, приведено в табл. 5.

Таблица 5

Table 5

Сравнительный анализ баз данных

Comparative database analysis

| База данных | Преимущества | Недостатки |
|-------------|---|--|
| PostgreSQL | Полностью совместима с SQL и обладает высокой масштабируемостью. Поддерживаются различные языки программирования, включая C/C++, Java, Perl, Python, Ruby и Tcl | При выполнении элементарных операций чтения PostgreSQL может показывать меньшую производительность |
| MySQL | Простота эксплуатации благодаря наличию графического интерфейса; высокая степень масштабируемости, позволяющая эффективно обрабатывать крупные объемы данных и сложные задачи | Некоторые операции в MySQL менее надежны. Хотя MySQL – проект с открытым исходным кодом, развитие продукта замедлилось, что может создать трудности при выявлении и устранении проблем |
| SQLite | Интуитивно понятный и простой в освоении SQL-интерфейс; минимальные требования к системным ресурсам и занимаемому месту; не нуждается в настройке или администрировании; база данных хранится в единственном файле, что упрощает ее перенос | Отсутствие удобного пользовательского интерфейса для управления; ограниченная возможность настройки из-за особенностей формата хранения данных |

После проведения сравнительного анализа баз данных была выбрана СУБД PostgreSQL, т. к. она совместима с SQL и обладает высокой масштабируемостью, а также поддерживает различные языки программирования.

Выбор языка программирования зависит от целей разработки: Python подходит для аналитики

и машинного обучения [10], Java широко используется в разработке любых приложений, R ориентирован на статистику и анализ данных, SQL необходим для работы с базами данных, а C++ применяется в системах с высокими требованиями к производительности (табл. 6).

Таблица 6

Table 6

Сравнительный анализ языков программирования для анализа данных

Comparative analysis of programming languages for data analysis

| Язык программирования | Скорость выполнения | Доступные платформы | Применение |
|-----------------------|---------------------|-----------------------|--|
| Python | Средняя | Windows, macOS, Linux | Аналитика, машинное обучение, веб-разработка |
| Java | Высокая | | Разработка любых приложений |
| R | Средняя | | Статистика, анализ данных |
| SQL | Высокая | | Работа с базами данных |
| C++ | | | Высокопроизводительные системы |

В качестве основного языка программирования для анализа данных был выбран Python.

Результаты сравнительного анализа средств визуализации представлены в табл. 7.

Таблица 7

Table 7

Сравнительный анализ средств визуализации
Comparative analysis of visualization tools

| Средство визуализации | Стоимость | Доступные платформы | Поддерживаемые языки |
|-----------------------------------|------------------------------------|-----------------------|--|
| Tableau | От 70 долл. в мес Бесплатно | Windows, macOS | Python (через API) |
| Power BI | | Windows | DAX, M, SQL |
| Google Data Studio | | Веб (браузер) | SQL, JavaScript (для кастомных отчетов) |
| Grafana | | Windows, macOS, Linux | SQL, PromQL, InfluxQL, JSON |
| Matplotlib (Python-библиотека) | | | Python |

Выбор инструмента визуализации зависит от потребностей системы. Tableau и Power BI предлагают мощные возможности визуализации, но требуют подписки. Grafana подходит для мониторинга данных в реальном времени. Google Data Studio является удобным веб-инструментом для отчетов. Matplotlib используется для встроенной аналитики в Python-приложениях.

В результате проведения сравнительного анализа был выбран инструмент Power BI, поскольку он предоставляет продвинутое встроенные сред-

ства аналитики и легко интегрируется с Python для вывода прогнозов.

Описание хранилища данных

Создание проекта архитектуры информационной системы прогнозирования цен на авиабилеты требует построения соответствующих диаграмм. Для определения ключевых сущностей базы данных спроектирована диаграмма классов информационной системы (рис. 3).

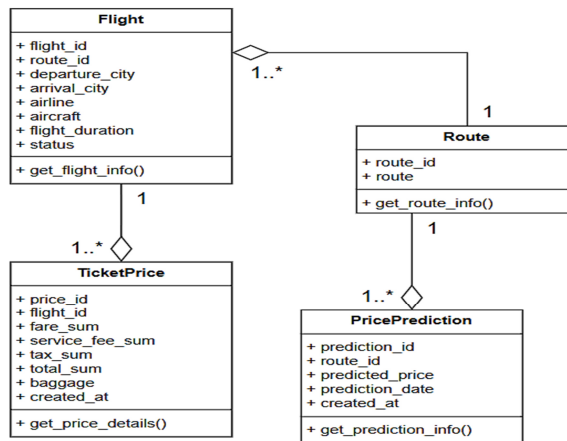


Рис. 3. Диаграмма классов информационной системы

Fig. 3. Information system class diagram

На диаграмме классов представлены:

1. Класс Flight (Рейс) предоставляет информацию о рейсе, включая его уникальный идентификатор, города отправления и назначения, тип самолета, продолжительность и статус рейса, а также он связан с классом Route через внешний ключ route_id. Метод get_flight_info() используется для получения информации о рейсе.

2. Класс TicketPrice (Цена на билет) связан с классом Flight через внешний ключ flight_id, что означает, что каждая запись о цене привязана к кон-

кретному рейсу. Атрибуты класса включают стоимость тарифа, сервисные сборы, налоги и общую стоимость билета. Метод get_price_details() агрегирует данные о базовом тарифе, сборах и налогах для получения итоговой стоимости конкретного билета.

3. Класс PricePrediction (Прогноз цены) связан с классом Flight через внешний ключ flight_id и с классом Route через внешний ключ route_id, хранит прогнозируемую цену и дату прогноза. Метод get_prediction_info() используется для получения информации о прогнозе.

4. Класс Route (Маршрут) предоставляет информацию о названиях маршрутов, хранит все возможные направления перелетов. Метод `get_route_info()` используется для получения информации о маршрутах.

На рис. 4 приведена схема базы данных, в которой учтены особенности СУБД PostgreSQL.

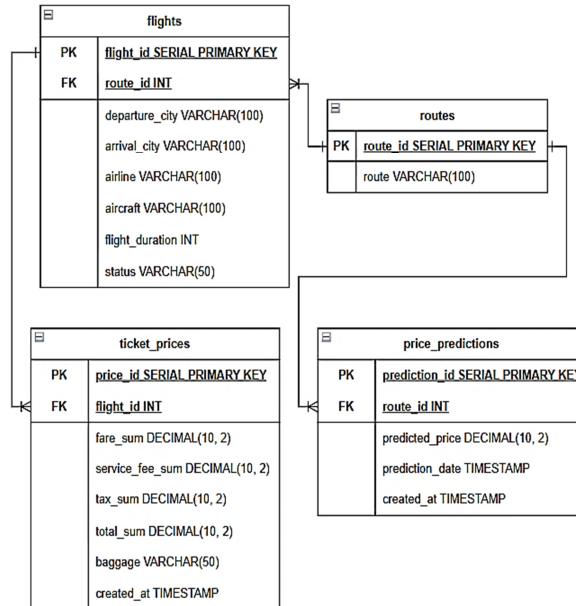


Рис. 4. Схема базы данных информационной системы

Fig. 4. Information system database schema

Структура приведена к третьей нормальной форме (ЗНФ) для устранения избыточности данных и оптимизации их обработки.

Описание потока данных

На рис. 5 для визуализации потоков данных произведено описание работы пользователей с разрабатываемой ИС на основе диаграммы последовательности.

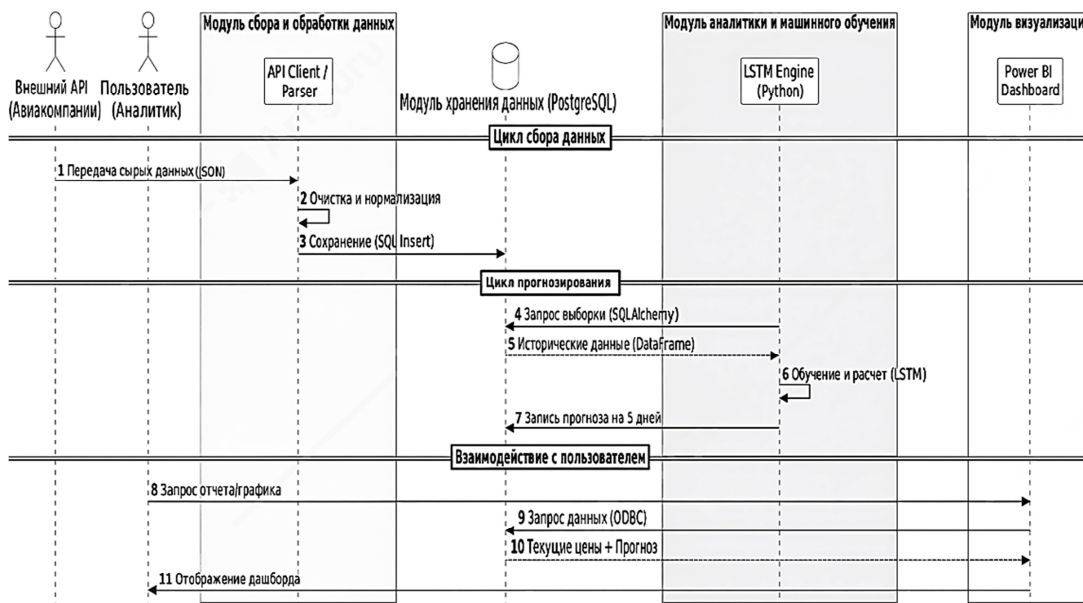


Рис. 5. Диаграмма последовательности информационной системы

Fig. 5. Information system sequence diagram

Реализация шагов диаграммы последовательности:

1. Пользователь (например, аналитик) вводит параметры для прогнозирования (например, город отправления, город назначения, дату вылета) и отправляет запрос на API-сервер.
2. API-сервер получает запрос от пользователя и передает его в систему прогнозирования.
3. Система прогнозирования (например, модуль на основе Visual Studio Code) обращается к базе данных (PostgreSQL) для получения данных о рейсах и ценах на билеты.
4. Модели машинного обучения обрабатывают исторические данные и прогнозируют будущее поведение цен на авиабилеты.

5. Прогнозируемые данные возвращаются в API-сервер, который передает результаты пользователю.
6. Пользователь получает прогнозы (например, в виде визуализаций с использованием Power BI). Это взаимодействие позволяет системе эффективно предсказывать цены на билеты, предоставляя пользователю необходимую информацию для принятия обоснованных решений.

Описание алгоритма прогнозирования

Диаграмма процесса разработанного алгоритма прогнозирования стоимости авиабилетов представлена на рис. 6.



Рис. 6. Диаграмма процесса информационной системы

Fig. 6. Information system process diagram

Поток данных включает этапы предварительной обработки, трансформацию временных рядов методом скользящего окна для обучения рекуррентной нейронной сети LSTM и последующую запись результатов в реляционную базу данных для сохранения прогнозов. Описание диаграммы процесса:

- необработанные данные преобразуются в очищенные. Происходит избавление от шума, дублей и пустых значений, чтобы модель не обучалась на ошибках;
- очищенные данные преобразуются в сгруппированные. Фокус исследования сужается: вместо всех рейсов алгоритм концентрируется на 10 самых

популярных маршрутах, т. к. по ним накоплено больше всего данных для обучения;

- числовые значения преобразуются в масштабированные в диапазоне от 0 до 1. Функции активации в LSTM чувствительны к масштабу, поэтому рубли переводятся в нормированный вид для стабильного обучения нейросети;
- временные ряды преобразуются в формат обучения с учителем. Данные собираются в окна. Модель анализирует показатели за последние дни, чтобы предсказать цену на следующий период;
- прогнозы преобразуются в реляционный SQL-формат. Результат прогноза объединяется с мета-

данными, такими как идентификатор маршрута и дата создания, и сохраняется в структурированном виде для дальнейшего использования, например для вывода на дашборд.

Анализ полученных результатов

С целью повышения эффективности работы информационной системы, ориентированной на прогнозирование цен на авиабилеты, важно определить ключевые метрики, которые будут измерять качество полученных прогнозов. Основной метрикой

является точность прогнозов (Prediction Accuracy) – мера того, насколько точно система предсказывает цены на авиабилеты. Это можно измерить с помощью метрик ошибок (например, RMSE, MAE). Метрики ошибок: RMSE (Root Mean Squared Error) – корень из среднеквадратичной ошибки, MAE (Mean Absolute Error) – средняя абсолютная ошибка.

Визуализация прогноза цен на авиабилеты в соответствующем модуле, программе Power BI, представлена на рис. 7.

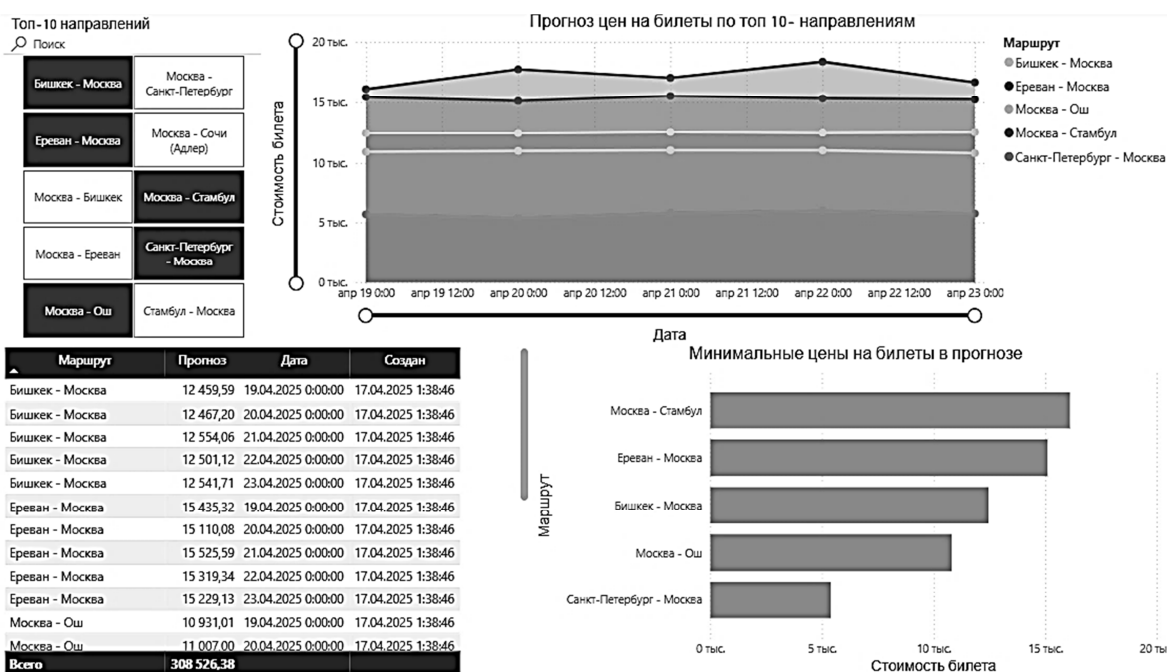


Рис. 7. Результат работы модуля визуализации

Fig. 7. The result of the visualization module operation

В качестве результата работы модуля прогнозирования были предсказаны цены на авиабилеты

по 5 самым популярным направлениям перелетов на ближайшие 5 календарных дней (табл. 8).

Таблица 8

Table 8

Результаты работы модуля прогнозирования

The results of the forecasting module

| Прогнозируемый маршрут | Название нормализованной метрики | Значение |
|--------------------------|----------------------------------|----------|
| Москва – Ереван | RMSE | 0,146 |
| | MAE | 0,117 |
| Москва – Бишкек | RMSE | 0,090 |
| | MAE | 0,072 |
| Москва – Стамбул | RMSE | 0,149 |
| | MAE | 0,115 |
| Москва – Ош | RMSE | 0,065 |
| | MAE | 0,049 |
| Санкт-Петербург – Москва | RMSE | 0,153 |
| | MAE | 0,112 |

С точки зрения работы прогнозирования в разработанной информационной системе, MAE показывает, насколько в среднем отличается прогноз от фактической цены. В данном случае значения в диапазоне 0,049–0,117, что говорит о достаточно точных прогнозах. RMSE слегка выше MAE (что ожидаемо, т. к. RMSE сильнее штрафует большие отклонения). Значения от 0,065 до 0,153 также считаются низкими, что говорит о стабильности модели и отсутствии крупных ошибок. Прогноз хорошо адаптирован для краткосрочного анализа (5 дней) – особенно важного при динамичном ценообразовании. Самые точные результаты – в 4-м тесте (RMSE = 0,065, MAE = 0,049) – можно использовать как эталон качества. Метрики показывают небольшие отклонения, и это означает, что предсказанные цены близки к реальным и могут быть основой для автоматизированных решений.

С точки зрения бизнеса точные прогнозы позволят выстроить эффективную ценовую стратегию и рекомендовать пользователям покупать билеты в оптимальный момент. Также информационная система позволяет авиакомпании повысить прозрачность ценообразования, что положительно сказывается на лояльности клиентов.

Основное преимущество внедрения системы для бизнеса – дополнительный доход, т. к. работа своевременных рекомендательных систем стимулирует продажи билетов на направления с ростом цен. Информационная система прогнозирования цен на авиабилеты показывает хорошую точность и стабильность, что делает ее ценным инструментом для бизнес-решений в авиакомпаниях.

Заключение

В условиях динамичного и конкурентного рынка авиаперевозок эффективное управление ценами на билеты становится ключевым фактором успешной работы компаний. Постоянные колебания стоимости, обусловленные множеством внешних и внутренних факторов, требуют не только оперативного реагирования, но и способности прогнозировать изменения цен. Технологии прогнозирования цен на авиабилеты

по заданным параметрам могут стать важным шагом к оптимизации процесса ценообразования.

Применение архитектуры с использованием Long Short-Term Memory (LSTM) и аналогичных моделей временных рядов выходит далеко за рамки прогнозирования цен на авиабилеты и демонстрирует высокую эффективность в различных отраслях, где ключевую роль играет анализ последовательных данных. Представленные результаты применимы в других областях. Способность LSTM выявлять скрытые закономерности в высоковолатильных данных позволяет масштабировать технологию на сферу финансов (прогнозирование биржевых курсов), энергетику (оптимизация работы энергосистем) и транспортную логистику (улучшение маршрутизации и управление складскими запасами).

В ходе исследования была разработана высокоточная информационная система для краткосрочного прогнозирования цен на авиабилеты. LSTM-модель показала лучшие результаты (MAPE = 4,2 %, $R^2 = 0,94$) по сравнению с Prophet, XGBoost и линейной регрессией. Также была реализована модульная архитектура: сбор данных с использованием PostgreSQL, аналитика при помощи Python, визуализация путем применения Power BI. Система способна нивелировать несовершенства рыночного анализа путем учета нелинейных колебаний спроса, устойчивости к выбросам и гибкости для интеграции с внешними данными.

В перспективе предполагается расширение функционала за счет добавления ансамблевых моделей (например, LSTM в сочетании с XGBoost) для комбинирования преимуществ подходов, а также внедрение онлайн-обучения для адаптации к резким изменениям цен (например, в период пандемии, кризиса) и масштабирование на другие отрасли: логистики (прогнозирование стоимости грузоперевозок), энергетики (предсказание цен на топливо), финансов (анализ волатильности акций транспортных компаний).

Предложенная система доказала свою эффективность в условиях высокой волатильности рынка авиаперевозок.

Список источников

1. Шедько Ю. Н., Власенко М. Н., Унижаев Н. В. Стратегическое управление проектами на основе использования искусственного интеллекта // Экономическая безопасность. 2021. Т. 4. № 3. С. 629–642.
2. Савин И. Н., Пастухова О. А. Использование методов прогнозирования для оптимизации ценообразования на авиабилеты // Наука и инновации в технике. 2023. № 5. С. 50–55.
3. Куликова Н. В., Ефимов И. А. Интеллектуальные технологии анализа ценовых данных на транспорте // Информационные технологии и телекоммуникации. 2022. № 3. С. 61–68.
4. Korkmaz H. Prediction of Airline Ticket Price Using Machine Learning Method // International Journal of Innovative Research in Science, Engineering and Technology. 2024. URL: <https://www.researchgate.net/publication/382607145> (дата обращения: 25.04.2025).
5. Кузьмина С. Н., Артемьев И. В. Методы машинного обучения в задачах динамического ценообразования // Вестн. Моск. ун-та. Сер. 15: Вычислительная математика и кибернетика. 2021. № 4. С. 78–84.
6. Плотников В. А., Колесникова Е. А. Прогнозирование временных рядов с помощью нейронных сетей LSTM // Вестн. Тамб. ун-та. Сер.: Естественные и технические науки. 2022. № 1. С. 122–127.
7. Белоусов Д. В., Ермаков А. И. Применение нейронных сетей LSTM для прогноза спроса в транспортной отрасли // Транспорт: наука, техника, управле-

ние. 2023. № 2. С. 45–50.

8. Pelka P., Dudek G. Pattern-based Long Short-term Memory for Mid-term Electrical Load Forecasting // arXiv preprint arXiv:2004.11834. 2020. URL: <https://arxiv.org/abs/2004.11834> (дата обращения: 25.04.2025).

9. Lara-Benítez P., Carranza-García M., Riquelme J. C. An Experimental Review on Deep Learning Architectures

for Time Series Forecasting // arXiv preprint arXiv:2103.12057. 2021. URL: <https://arxiv.org/abs/2103.12057> (дата обращения: 25.04.2025).

10. Голубев П. М., Чернышев А. В. Анализ и прогнозирование временных рядов в Python: практическое руководство // Современные технологии в науке и образовании. 2020. № 6. С. 33–37.

References

1. Shed'ko Yu. N., Vlasenko M. N., Unizhaev N. V. Strategicheskoe upravlenie proektami na osnove ispol'zovaniya iskusstvennogo intellekta [Strategic project management based on the use of artificial intelligence]. *Ekonomicheskaya bezopasnost'*, 2021, vol. 4, no. 3, pp. 629-642.

2. Savin I. N., Pastuhova O. A. Ispol'zovanie metodov prognozirovaniya dlya optimizatsii cenoobrazovaniya na aviabiletu [Using forecasting methods to optimize airline ticket pricing]. *Nauka i innovatsii v tekhnike*, 2023, no. 5, pp. 50-55.

3. Kulikova N. V., Efimov I. A. Intellektual'nye tekhnologii analiza cenovykh dannykh na transporte [Intelligent technologies for analyzing price data in transport]. *Informacionnye tekhnologii i telekommunikatsii*, 2022, no. 3, pp. 61-68.

4. Korkmaz H. Prediction of Airline Ticket Price Using Machine Learning Method. *International Journal of Innovative Research in Science, Engineering and Technology*, 2024. Available at: <https://www.researchgate.net/publication/382607145> (accessed: 25.04.2025).

5. Kuz'mina S. N., Artem'ev I. V. Metody mashinnogo obucheniya v zadachah dinamicheskogo cenoobrazovaniya [Machine learning methods in dynamic pricing tasks]. *Vestnik Moskovskogo universiteta. Seriya 15: Vychislitel'naya matematika i kibernetika*, 2021, no. 4, pp. 78-84.

6. Plotnikov V. A., Kolesnikova E. A. Prognozirovanie

vremennykh ryadov s pomoshch'yu nejronnykh setej LSTM [Prediction of time series using LSTM neural networks]. *Vestnik Tambovskogo universiteta. Seriya: Estestvennye i tekhnicheskie nauki*, 2022, no. 1, pp. 122-127.

7. Belousov D. V., Ermakov A. I. Primenenie nejronnykh setej LSTM dlya prognoza sprosa v transportnoj otrasli [Application of LSTM neural networks for forecasting demand in the transportation industry]. *Transport: nauka, tekhnika, upravlenie*, 2023, no. 2, pp. 45-50.

8. Pelka P., Dudek G. Pattern-based Long Short-term Memory for Mid-term Electrical Load Forecasting. *arXiv preprint arXiv:2004.11834*, 2020. Available at: <https://arxiv.org/abs/2004.11834> (accessed: 25.04.2025).

9. Lara-Benítez P., Carranza-García M., Riquelme J. C. An Experimental Review on Deep Learning Architectures for Time Series Forecasting. *arXiv preprint arXiv:2103.12057*, 2021. Available at: <https://arxiv.org/abs/2103.12057> (accessed: 25.04.2025).

10. Golubev P. M., Chernyshev A. V. Analiz i prognozirovanie vremennykh ryadov v Python: prakticheskoe rukovodstvo [Time Series Analysis and Forecasting in Python: a practical guide]. *Sovremennye tekhnologii v nauke i obrazovanii*, 2020, no. 6, pp. 33-37.

Статья поступила в редакцию 12.05.2025; одобрена после рецензирования 19.01.2026; принята к публикации 15.04.2026
The article was submitted 12.05.2025; approved after reviewing 19.01.2026; accepted for publication 15.04.2026

Информация об авторах / Information about the authors

Татьяна Евгеньевна Саратова – доктор технических наук, доцент; заведующий кафедрой прикладной математики; МИРЭА – Российский технологический университет; saratova@mirea.ru

Tatyana E. Saratova – Doctor of Technical Sciences, Assistant Professor; Head of the Department of Applied Mathematics; MIREA – Russian Technological University; saratova@mirea.ru

Илья Сергеевич Кожевников – генеральный директор; ООО «Айти класс»; is.kozhevnikov@gmail.com

Ilya S. Kozhevnikov – Chief Executive Officer; “IT-Class” LLC; is.kozhevnikov@gmail.com

Владислав Андреевич Шульгин – студент направления «Информационная безопасность (Киберразведка и противодействие угрозам с применением технологий искусственного интеллекта)»; МИРЭА – Российский технологический университет; vladshulgin9@gmail.com

Vladislav A. Shulgin – Student of the course “Information security (Cyber intelligence and counteraction to threats using artificial intelligence technologies)”; MIREA – Russian Technological University; vladshulgin9@gmail.com

