DOI: 10.12731/2227-930X-2025-15-2-347 **EDN: HHJBXM**
УДК 004.032.26

CC BY NC ND

Научная статья | Транспортные и транспортно-технологические системы

ПРОГНОЗИРОВАНИЕ ТРАНСПОРТНОЙ ЗАГРУЖЕННОСТИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Р.Р. Загидуллин, А.Н. Хайбуллин

Аннотация

Обоснование. В статье разрабатывается комплексный подход к прогнозированию транспортной загруженности с использованием синтетических данных, имитирующих динамику городского трафика. Гибридная методология позволяет объединить анализ временных рядов и глубокое обучение, что актуально для моделирования нелинейных зависимостей и закономерностей в транспортных данных.

Цель. Целью работы является разработка и тестирование прогностической модели, способной точно предсказывать уровни транспортной загруженности с учётом сезонных и погодных факторов.

Материалы и методы. Для выявления паттернов в данных применено аддитивное разложение временного ряда, спектральный анализ на основе быстрого преобразования Фурье и оценка автокорреляционных зависимостей. Прогностическая модель реализована в виде двухэтапного подхода: классический алгоритм ARIMA используется для базового прогнозирования, а архитектура LSTM с двумя рекуррентными слоями и регуляризацией – для обучения на последовательностях длиной 24 часа. Дополнительно для сопоставления и подтверждения результатов применён ансамблевый метод Random Forest, настроенный с гиперпараметрами: 200 деревьев, максимальная глубина – 12, минимальное количество объектов в листе – 2.

Результаты. Результаты демонстрируют превосходство LSTM-модели над ARIMA и Random Forest по точности предсказаний, что подтверждается визуальным сопоставлением прогнозов с тестовыми данными и метрикой среднеквадратичной ошибки. Выявлены ключевые факторы, влияющие на загруженность: суточные циклы интенсивности трафика, рост нагрузки при осадках (до 30% при снеге и 20% при дожде), а также температурно-зависимая модуляция транспортного потока.

Ключевые слова: транспортная загруженность; транспортное средство; нейронная сеть; машинное обучение; прогнозирование

Для цитирования. Загидуллин, Р. Р., & Хайбуллин, А. Н. (2025). Прогнозирование транспортной загруженности с использованием методов машинного обучения. *International Journal of Advanced Studies*, *15*(2), 202–216. https://doi.org/10.12731/2227-930X-2025-15-2-347

Original article | Transport and Transport-Technological Systems

TRAFFIC CONGESTION FORECASTING USING MACHINE LEARNING METHODS

R.R. Zagidullin, A.N. Khaybullin

Abstract

Background. This study develops a comprehensive approach to traffic congestion forecasting using synthetic data that simulates the dynamics of urban traffic. A hybrid methodology is proposed that combines time series analysis and deep learning, which is highly relevant for modeling nonlinear dependencies and patterns in traffic data.

Purpose. The purpose of this study is to develop and test a predictive model capable of accurately forecasting traffic congestion levels while accounting for seasonal and weather-related factors.

Materials and methods. To identify patterns in the data, additive time series decomposition, spectral analysis based on the fast Fourier transform,

and autocorrelation analysis were applied. The predictive model was implemented using a two-stage approach: the classical ARIMA algorithm was used for baseline forecasting, while an LSTM architecture with two recurrent layers and regularization was trained on 24-hour sequences. Additionally, to compare and validate the results, the ensemble method Random Forest was used, configured with the following hyperparameters: 200 trees, maximum depth of 12, minimum samples per leaf of 2.

Results. The results demonstrate the superiority of the LSTM model over ARIMA and Random Forest in terms of predictive accuracy, as confirmed by visual comparison of forecasts with test data and by the mean squared error metric. Key factors influencing congestion were identified, including daily traffic intensity cycles, increased load during precipitation events (up to 30% during snow and 20% during rain), as well as temperature-dependent modulation of traffic flow.

Keywords: traffic congestion; vehicle; neural network; machine learning; forecasting

For citation. Zagidullin, R. R., & Khaybullin, A. N. (2025). Traffic congestion forecasting using machine learning methods. *International Journal of Advanced Studies*, *15*(2), 202–216. https://doi.org/10.12731/2227-930X-2025-15-2-347

Введение

Многие населенные пункты сталкиваются с растущими вызовами в организации транспортных потоков, где традиционные подходы к управлению инфраструктурой и логистикой [4-6] становятся недостаточно эффективными. В этом контексте интеллектуальные транспортные системы (ИТС), объединяющие IoT-сенсоры, облачные вычисления и алгоритмы анализа данных, формируют новый парадигмальный подход к оптимизации дорожных сетей. Ключевым элементом таких систем выступает прогнозирование транспортной загруженности [7; 8], позволяющее реагировать на текущие изменения трафика и осуществлять предиктивное управление маршрутизацией.

Актуальность задачи обусловлена нелинейной зависимостью интенсивности движения от множества факторов: от циклических колебаний (суточных, недельных, сезонных) до различных переменных, таких как погодные аномалии. Современные исследования подтверждают, что даже кратковременные метеоусловия могут приводить к статистически значимым изменениям трафика: так, среднесуточное увеличение интенсивности на 20–30% наблюдается при неблагоприятной погоде в утренние часы пик. Более того, в условиях мегаполисов эффект резонанса между осадками и транспортной плотностью приводит к существенному росту задержек и снижению средней скорости потока.

Особенности исследования и данные

Исследование подобных задач требует разработки гибридных моделей, способных интегрировать гетерогенные данные в единый прогнозный контур. В частности, важной задачей становится моделирование взаимодействия между метеофакторами, социальными событиями (праздничные дни, массовые мероприятия) и типичными шаблонами трафика. Результаты эмпирических работ показывают, что игнорирование хотя бы одного из этих факторов приводит к смещению прогноза на интервале 5–15%, что в практическом плане может выражаться в необоснованной перегрузке отдельных магистралей.

В рамках настоящего исследования фокус смещен на исследование методов машинного обучения [9; 10] и анализа временных рядов для создания адаптивного инструмента прогнозирования транспортного движения. Особенностью подхода является использование синтетически сгенерированных данных, имитирующих реальные условия населенного пункта с контролируемыми параметрами:

- сезонные колебания базовой нагрузки,
- нелинейное влияние осадков (дождь/снег),
- температурно-зависимая модификация трафика.

Такой подход позволяет валидировать модели в условиях ограниченности реальных данных и проводить стресс-тестирование для экстремальных сценариев. В ходе численных экспериментов была оценена устойчивость модели к варьированию погодных параметров и цикличности: выявлено, что исключение суточных паттернов из обучающего множества снижает точность прогноза на 25–27%, в то время как исключение погодных данных – на 10–15%, что демонстрирует относительный вклад каждого из факторов.

Методы прогнозирования транспортной загруженности

Основной научный вклад работы заключается в комбинации методов частотного анализа (FFT) для детекции скрытых периодичностей с архитектурами LSTM, ориентированными на долгосрочные зависимости, – синтез, редко встречающийся в современных исследованиях транспортных систем. Практическая реализация разработанной модели может включать её интеграцию в городские центры управления движением для автоматической адаптации сигналов светофоров и прогнозирования загрузки ключевых магистралей. Такой подход позволит повысить пропускную способность транспортной сети и снизить задержки на дорогах за счёт адаптивного управления трафиком в режиме реального времени.

Для вычислений созданы синтетические данные, имитирующие поведение транспортного потока. В модели учитывается влияние температуры и погодных условий на интенсивность движения. Формально индекс загруженности определяется по следующей формуле:

$$C_t = \frac{T(t)}{\Theta(t)+1} * \left(1 + \beta_{snow} * 1_{P(t)=snow} + \beta_{rain} * 1_{P(t)=rain}\right), \quad (1)$$

где:

- $T(t)$ – интенсивность трафика

$$T(t) = T_0 + A * \sin\left(\frac{2\pi t}{24}\right) + \varepsilon_t, \quad (2)$$

- θ(t) – параметр, который отражает влияние температуры,
- β – случайный шум,

- $1_{P(t)=snow}$, $1_{P(t)=rain}$ – индикаторы наличия снега и дождя.

Случайный шум добавляется для имитации реальных данных.

Для выявления трендов и сезонных колебаний применяется аддитивное разложение временного ряда по формуле:

$$X(t) = T(t) + S(t) + R(t), \quad (3)$$

где:

- $T(t)$ – трендовая составляющая,
- $S(t)$ – сезонная компонента,
- $R(t)$ – случайная (шумовая) компонента.

Этот метод позволяет визуально оценить характер суточных и годовых циклов в данных.

Чтобы выявить частотные компоненты, применяется быстрое преобразование Фурье (FFT). Функция Фурье описывается следующим выражением:

$$F(\omega) = \sum_{t=0}^{N-1} x_t e^{-\frac{2\pi i \omega t}{N}}, \quad (4)$$

Анализ спектра позволяет обнаружить доминирующие циклы, что важно для прогноза сезонных изменений в трафике.

Модель LSTM

Для прогнозирования временных рядов используется рекуррентная нейронная сеть [11, 12] LSTM. Перед подачей данных в модель происходит нормализация по следующей формуле:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}. \quad (5)$$

Затем данные преобразуются в последовательности длиной 24 часа, что соответствует одному дню. Базовая архитектура модели [13; 14] включает два слоя LSTM, между которыми применяется слой Dropout для предотвращения переобучения, и финальный полносвязный слой с линейной активацией:

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b), \quad (6)$$

$$y_t = W_o h_t + b_o, \quad (7)$$

где h_t – скрытое состояние, а W и b – веса и смещения соответственно.

Для обучения модели использован алгоритм оптимизации Adam с настройкой весов на минимизацию среднеквадратичной ошибки (MSE):

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2, \quad (8)$$

Adam адаптивно корректирует скорость обучения, что важно для сходимости модели при работе с нестационарными временными рядами, а MSE обеспечивает чёткую оценку отклонений прогноза от реальных значений. Дополнительно использование Adam ускоряет процесс обучения [15; 16] за счёт автоматической настройки шага градиентного спуска.

Применим описанную модель для прогнозирования интенсивности трафика и сравним полученные результаты с графиком фактических данных.

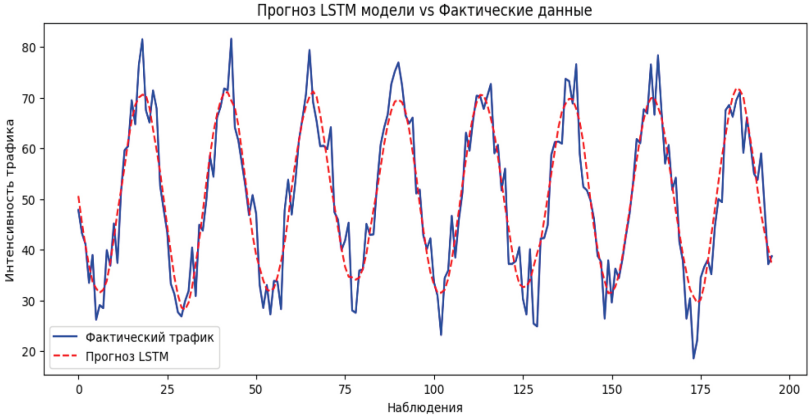

Рис. 1. Графики интенсивности трафика: линии фактических данных и прогнозов модели LSTM

Величина среднеквадратической ошибки (MSE) для модели LSTM = 2.19.

Модель ARIMA

Для комплексного анализа временных зависимостей в данных транспортного трафика применена модель ARIMA (Autoregressive Integrated Moving Average). Её ключевое преимущество заключается в способности описывать стационарные и нестационарные ряды через комбинацию авторегрессии (AR), интегрирования (I) и скользящего среднего (MA). Такой подход позволяет учесть как краткосрочные автокорреляции (через AR), так и долгосрочные тренды и сезонность (через I и MA), обеспечивая гибкость при работе с данными, в которых присутствуют циклические и сезонные колебания транспортного потока. Кроме того, ARIMA модель обладает хорошо разработанным математическим аппаратом для оценки остатков, автокорреляций и ошибок прогноза, что позволяет не только строить прогнозы, но и проверять их адекватность с использованием статистических тестов (например, теста Льюнг-Бокса или анализа автокорреляционной функции остатков).

Для оценки зависимости текущих значений трафика от предыдущих использована автокорреляционная (ACF) функция:

$$p_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}, \quad (9)$$

Эта функция помогает определить порядок модели ARIMA.

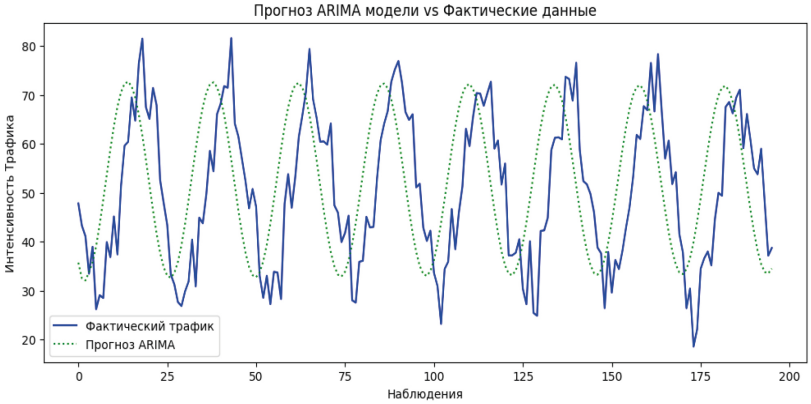

Рис. 2. Графики интенсивности трафика: линии фактических данных и прогнозов модели ARIMA

Применим описанную модель для прогнозирования интенсивности трафика и сравним полученные результаты с графиком фактических данных.

Величина среднеквадратической ошибки (MSE) для модели ARIMA = 3.08.

Модель Random Forest

Метод Random Forest представляет собой ансамблевый алгоритм машинного обучения [17], основанный на объединении большого количества решающих деревьев (Decision Trees). В контексте прогнозирования транспортной загруженности он используется для предсказания значения трафика (например, интенсивности или индекса загруженности) на основе исторических данных. Random Forest использует бэггинг (bagging, bootstrap aggregating), то есть создаёт множество деревьев решений на разных подвыборках обучающей выборки и усредняет их прогнозы, чтобы уменьшить переобучение и повысить точность модели.

Применим описанную модель для прогнозирования интенсивности трафика и сравним полученные результаты с графиком фактических данных.

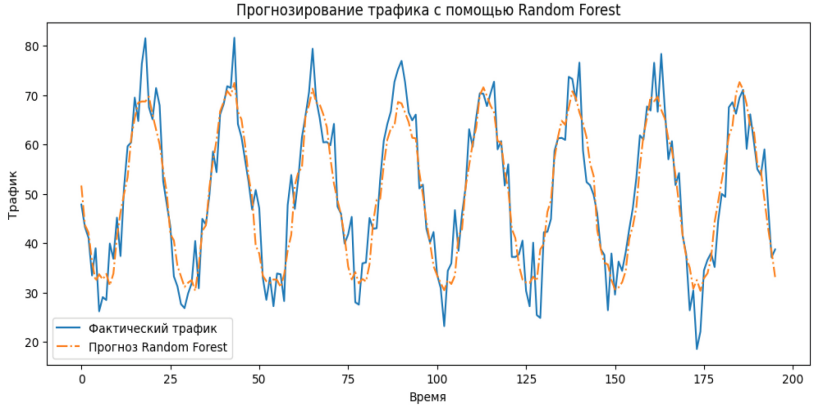

Рис. 3. Графики интенсивности трафика: линии фактических данных и прогнозов модели Random Forest

Величина среднеквадратической ошибки (MSE) для модели Random Forest равна 5.29.

Результаты

В данной задаче прогнозирования транспортной загруженности применялись три модели: Random Forest, ARIMA и LSTM.

ARIMA эффективно справилась с описанием сезонности и трендов в данных. Она показала хорошую точность (MSE = 3.08) и позволила учесть как краткосрочные автокорреляции, так и долгосрочные изменения транспортного потока. Применение статистических тестов (например, теста Льюнг-Бокса) подтвердило адекватность модели и её способность предсказывать временные зависимости.

LSTM, благодаря своей архитектуре, способной улавливать долгосрочные зависимости, показала наилучшие результаты среди всех моделей (MSE = 2.19). Использование нормализации данных и архитектуры с двумя слоями LSTM, а также алгоритма оптимизации Adam, позволило минимизировать ошибку прогноза и более точно учитывать циклические изменения интенсивности движения.

Random Forest показал, что настройка гиперпараметров (n_estimators, max_depth, min_samples_leaf, max_features) существенно влияет на стабильность и точность прогноза. Для повышения устойчивости при наличии шума в данных и учёте циклических колебаний трафика использовались параметры: n_estimators = 200, max_depth = 12, min_samples_leaf = 2 и max_features = sqrt(N). Несмотря на то что Random Forest продемонстрировал менее высокую точность (MAE = 4.20, MSE = 5.29) по сравнению с другими методами, он обеспечил полезную базовую линию для оценки прогноза.

Таким образом, наилучшие результаты были достигнуты с помощью LSTM, а ARIMA показала достойную точность в учёте сезонности и трендов. Random Forest, несмотря на ограниченность в учёте порядка временных наблюдений, дал ценную базовую оценку и помог валидации других моделей.

Заключение

Разработана гибридная модель прогнозирования транспортной загруженности, сочетающая методы анализа временных рядов (ARIMA) и глубокого обучения (LSTM). Результаты экспериментов подтверждают эффективность предложенного подхода: модель LSTM демонстрирует наилучшую точность прогнозирования с MSE = 2.19, что свидетельствует о её способности учитывать нелинейные зависимости в данных. ARIMA показала себя достойно, хотя её точность (MSE = 3.08) несколько уступает. Модель Random Forest имела наибольшую погрешность (MSE = 5.29), что указывает на ограниченную способность Random Forest адекватно моделировать сложные временные зависимости в данных о транспортной загруженности.

Визуальный анализ предсказаний выявил хорошее соответствие прогнозных значений тестовым данным, особенно для суточных циклов интенсивности трафика. Проведённые вычисления подтверждают практическую применимость модели для краткосрочного прогнозирования.

Перспективы работы связаны с интеграцией дополнительных параметров (например, данных о событиях на дорогах), использованием ансамблевых методов и тестированием на реальных данных городской инфраструктуры. Использование подобных моделей в системах управления движением также способствует снижению выбросов CO_2 за счёт уменьшения времени простоя автомобилей в пробках. Это не только улучшает экологическую обстановку в городах, но и способствует улучшению качества жизни населения благодаря сокращению времени в пути и повышению предсказуемости поездок.

Список литературы

1. Хамидулин, Т. Г. (2019). Применение искусственных нейронных сетей в транспортной отрасли. *Экономика и социум*, *(4)*, 851–858. EDN: https://elibrary.ru/ZZRFOP

2. Галушкин, А. И. (2015). *Нейронные сети: основы теории*. Москва: РиС. 496 с.
3. Редько, В. Г. (2019). *Эволюция, нейронные сети, интеллект: модели и концепции эволюционной кибернетики*. Москва: Ленанд. 224 с.
4. Бочкарев, А. А., & Бочкарев, П. А. (2024). *Логистика городских транспортных систем: учебное пособие для вузов* (3-е изд., перераб. и доп.). Москва: Юрайт. 162 с. EDN: https://elibrary.ru/BVRIOQ
5. Герами, В. Д., & Колик, А. В. (2024). *Управление транспортными системами. Транспортное обеспечение логистики: учебник и практикум для вузов* (3-е изд., перераб. и доп.). Москва: Юрайт. 536 с. EDN: https://elibrary.ru/BCZLEH
6. Андреева, Л. А., & др. (2015). *Инновационные процессы логистического менеджмента в интеллектуальных транспортных системах* (под ред. Миротина, Л. Б., & Левина, Б. А.). Том 2. Формирование отраслевых логистических интеллектуальных транспортных систем. Москва: Учебно-методический центр по образованию на железнодорожном транспорте. 343 с.
7. Душкин, Р. В. (2020). Интеллектуальные транспортные системы: моделирование и прогнозирование трафика. *Транспортные системы и технологии*, *(4)*, 45–62.
8. Boukerche, A., & Wang, J. (2020). Machine learning-based traffic prediction models for intelligent transportation systems. *Computer Networks*, *181*, 107530. https://doi.org/10.1016/j.comnet.2020.107530 EDN: https://elibrary.ru/HKZXAK
9. Николенко, С. И., Кадурин, А., & Архангельская, Е. В. (2018). *Глубокое обучение. Погружение в мир нейронных сетей*. Санкт-Петербург: Питер. 480 с.
10. Флах, П. (2015). *Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных*. Москва: ДМК Пресс. 400 с.
11. Бурков, А. (2020). *Машинное обучение без лишних слов*. Москва: Альпина Паблишер. 100 с.

12. Фаустова, К. И. (2017). Нейронные сети: применение сегодня и перспективы развития. *Территория науки*, *(4)*, 83–87. EDN: https://elibrary.ru/ZXPNRL
13. Иванько, А. Ф., Иванько, М. А., & Сизова, Ю. А. (2019). Нейронные сети: общие технологические характеристики. *Научное обозрение. Технические науки*, *(2)*, 17–23. EDN: https://elibrary.ru/QHUGLR
14. Хайкин, С. (2006). *Нейронные сети: полный курс* (2-е изд.). Москва: Вильямс. 1104 с.
15. Ежов, А. А., & Шумский, С. А. (1998). *Нейрокомпьютинг и его применения в экономике и бизнесе*. Москва: МИФИ. 268 с.
16. Тархов, Д. А. (2005). *Нейронные сети. Модели и алгоритмы*. Москва: Радиотехника.
17. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York: Springer-Verlag.

References

1. Khamidulin, T. G. (2019). Application of artificial neural networks in the transport industry. *Economics and Society*, *(4)*, 851–858. EDN: https://elibrary.ru/ZZRFOP
2. Galushkin, A. I. (2015). *Neural networks: Fundamentals of theory*. Moscow: RIS. 496 p.
3. Red'ko, V. G. (2019). *Evolution, neural networks, intelligence: Models and concepts of evolutionary cybernetics*. Moscow: Lenand. 224 p.
4. Bochkarev, A. A., & Bochkarev, P. A. (2024). *Urban transport logistics systems: A textbook for universities* (3rd ed., revised and supplemented). Moscow: Yurayt. 162 p. EDN: https://elibrary.ru/BVRIOQ
5. Gerami, V. D., & Kolik, A. V. (2024). *Transport systems management. Transport logistics support: A textbook and practical guide for universities* (3rd ed., revised and supplemented). Moscow: Yurayt. 536 p. EDN: https://elibrary.ru/BCZLEH
6. Andreeva, L. A., et al. (2015). *Innovative processes of logistics management in intelligent transport systems* (under the editorship of Miro-

tin, L. B., & Levin, B. A.). Volume 2. Formation of industry-specific intelligent transport logistics systems. Moscow: Educational and Methodological Center for Education in Railway Transport. 343 p.

7. Dushkin, R. V. (2020). Intelligent transport systems: Traffic modeling and forecasting. *Transport Systems and Technologies*, *(4)*, 45–62.
8. Boukerche, A., & Wang, J. (2020). Machine learning-based traffic prediction models for intelligent transportation systems. *Computer Networks*, *181*, 107530. https://doi.org/10.1016/j.comnet.2020.107530 EDN: https://elibrary.ru/HKZXAK
9. Nikolaenko, S. I., Kadurin, A., & Arkhangelskaya, E. V. (2018). *Deep learning: Diving into the world of neural networks*. Saint Petersburg: Piter. 480 p.
10. Flach, P. (2015). *Machine learning: The science and art of building algorithms that extract knowledge from data*. Moscow: DMK Press. 400 p.
11. Burkov, A. (2020). *Machine learning without unnecessary words*. Moscow: Alpina Publisher. 100 p.
12. Faustova, K. I. (2017). Neural networks: Current applications and future development prospects. *Territory of Science*, *(4)*, 83–87. EDN: https://elibrary.ru/ZXPNRL
13. Ivanko, A. F., Ivanko, M. A., & Sizova, Yu. A. (2019). Neural networks: General technological characteristics. *Scientific Review. Engineering Sciences*, *(2)*, 17–23. EDN: https://elibrary.ru/QHUGLR
14. Haykin, S. (2006). *Neural networks: A comprehensive foundation* (2nd ed.). Moscow: Williams. 1104 p.
15. Ezhov, A. A., & Shumsky, S. A. (1998). *Neurocomputing and its applications in economics and business*. Moscow: MEPhI. 268 p.
16. Tarkhov, D. A. (2005). *Neural networks. Models and algorithms*. Moscow: Radiotekhnika.
17. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York: Springer-Verlag.

ДАННЫЕ ОБ АВТОРАХ

Загидуллин Рамиль Равильевич, кандидат технических наук, доцент кафедры конструктивно-дизайнерского проектирования Института дизайна и пространственных искусств
Казанский (Приволжский) федеральный университет
ул. Кремлевская, 18, г. Казань, 420008, Российская Федерация
Ramil.Zagidullin@kpfu.ru

Хайбуллин Алмаз Наилевич, техник НИЛ «Интеллектуальная мобильность» Института дизайна и пространственных искусств
Казанский (Приволжский) федеральный университет
ул. Кремлевская, 18, г. Казань, 420008, Российская Федерация
almaz.khaybullin@mail.ru

DATA ABOUT THE AUTHORS

Ramil R. Zagidullin, Candidate of Technical Sciences, Associate Professor of the Department of Structural and Design Engineering, Institute of Design and Spatial Arts
Kazan (Volga region) Federal University
18, Kremlevskaya Str., Kazan, 420008, Russian Federation
Ramil.Zagidullin@kpfu.ru
ORCID: https://orcid.org/0000-0001-5185-2690
ResearcherID: E-5671-2018
Scopus ID: 57193743308

Almaz N. Khaybullin, Technician at the "Intelligent Mobility" Research Laboratory (R&D Lab), Institute of Design and Spatial Arts
Kazan (Volga region) Federal University
18, Kremlevskaya Str., Kazan, 420008, Russian Federation
almaz.khaybullin@mail.ru
ORCID: https://orcid.org/0009-0001-7321-956X
ResearcherID: LMP-8981-2024
Scopus ID: 5830790020

Поступила 05.06.2025
После рецензирования 18.06.2025
Принята 23.06.2025

Received 05.06.2025
Revised 18.06.2025
Accepted 23.06.2025