

# Прогнозирование срывов сроков строительства с использованием машинного обучения на основе исторических данных о фактической продолжительности завершённых проектов

В.В. Коньков, В.И. Широков, М.Г. Жабицкий

**Аннотация** — В настоящем исследовании рассматривается проблема превышения плановых сроков выполнения строительных работ, являющаяся существенным препятствием для повышения эффективности строительной отрасли и её вклада в экономику Российской Федерации. В рамках задачи, определённой стратегией развития строительной отрасли и жилищно-коммунального хозяйства России на период до 2030 года, предусмотрено сокращение продолжительности инвестиционно-строительного цикла на 30%. Тем не менее, действующие методы планирования, основанные на устаревших нормативных подходах, показали свою неэффективность вследствие невнимания к статистическим данным уже завершённых проектов.

Для решения данной проблемы авторами предложена гипотеза о создании системы рекомендаций на основе анализа исторических данных выполнения отдельных строительных работ, строительных проектов. В качестве исходных данных использовались обезличенные сведения о плановых и фактических сроках выполнения работ, обработанные методами разведочного анализа данных и машинного обучения. Полученные результаты позволили выявить закономерности в реализации строительных проектов и разработать методику прогнозирования срывов сроков отдельных работ, направленную на оптимизацию планирования и сокращение длительности планируемых к реализации строительных проектов.

Внедрение предложенного подхода и его апробация на реальных строительных проектах обеспечивают руководителям своевременную информацию для корректировки графиков и повышения эффективности управления проектами.

**Ключевые слова** — Задержки в строительстве, машинное обучение, прогнозное моделирование, анализ исторических данных, управление проектами.

## I. ВВЕДЕНИЕ

В настоящее время строительная отрасль сталкивается с рядом вызовов, преодоление которых может значительно повысить её эффективность и укрепить вклад в экономику страны. Согласно стратегии развития строительной отрасли и жилищно-коммунального хозяйства Российской Федерации на период до 2030 года [1], одной из ключевых задач является сокращение продолжительности инвестиционно-строительного

цикла как минимум на 30 процентов [2]. Однако достижение этой цели осложняется тем, что фактические сроки выполнения работ на объектах превышают плановые.

Для того, чтобы обеспечить руководителей строительных проектов своевременной информацией о возможных изменениях в графике на этапе планирования и устранить задержки в строительных проектах, где отклонения от графика часто превышают 50 % от запланированного, предложена гипотеза о возможности создания системы рекомендаций на основе исторических данных. В отличие от существующих методов, которые опираются на устаревшие нормативные подходы и не учитывают статистические данные по реализованным проектам [3], данная система будет учитывать реальные исторические данные для более точного прогнозирования сроков и выявления потенциальных задержек.

Наиболее важной ценностью данного исследования является демонстрация того, как можно показать руководителям проектов фактическую продолжительность работы над аналогичными завершёнными проектами, используя рекомендательную систему на основе исторических данных, что позволит предоставить руководителям проектов заблаговременную информацию об изменениях графика на этапе планирования и исключить задержки в строительных проектах.

Цель данной статьи заключается в разработке и расширении текущих знаний посредством создания рекомендательной системы на основе машинного обучения для предсказания задержек в строительных проектах, использующих исторические данные. Мы стремимся оптимизировать процессы принятия решений руководителями проектов, предоставляя рекомендации по работам графика в виде прогнозной даты завершения. Используя методы машинного обучения, в частности обучения с учителем, была обучена модель на предварительно размеченном наборе данных из 4000 наименований строительных работ. Для классификации пользовательских наименований строительных работ, что позволило создать рабочий график, указывающий ожидаемую продолжительность работ и выделяющий отклонения от плановых сроков выполнения. Наши

Статья получена 26 июня 2024.  
Коньков Владислав Владимирович, аспирант ИИКС НИЯУ МИФИ (e-mail: vlad.konkov.7145@gmail.com)  
Широков Валерий Игоревич, магистр ВИШ НИЯУ МИФИ (e-mail: valerkashir@gmail.com)  
Жабицкий Михаил Георгиевич, заместитель директора ВИШ НИЯУ МИФИ (e-mail: jabitsky@mail.ru)

основные цели состоят в минимизации задержек и повышении точности и своевременности прогнозов графиков, что способствует более надежному управлению проектами.

Разработанная рекомендательная система, основанная на исторических данных, позволяет руководителям проектов получать оценочные данные о продолжительности работ на основе их названия и размера проекта. Эта модель предоставляет ценные инсайты в отношении прогнозируемого графика для всего проекта в сравнении с целевыми датами, облегчая проактивное принятие решений и смягчение потенциальных задержек.

Статья структурирована следующим образом: Во введении мы излагаем теоретическую гипотезу, составляющую основу нашей системы рекомендаций. В разделе 2 приводится литературный обзор, детализирующий современное состояние исследований и схожие методы решения. В разделе 3 подробно описана методология исследования, в разделе 4 указаны источники данных, а в разделе 5 продемонстрированы результаты предварительной обработки и фильтрации данных. В разделе 6 продемонстрированы результаты анализа распределений работ и выявлена необходимость автоматической классификации пользовательских наименований работ средствами машинного обучения. В разделе 8 обсуждаются возможность построения прогнозных кривых и приводятся результаты подтверждения гипотезы. Наконец, статья завершается в разделе 9, в котором обобщаются ключевые идеи и полученные результаты, а в разделе 10 предлагаются перспективы и направления будущих исследований в области применения машинного обучения в управлении строительными проектами.

## II. ЛИТЕРАТУРНЫЙ ОБЗОР

Современное состояние дел в области машинного обучения в строительстве фокусируется на прогнозировании затрат и оптимизации графиков. Несмотря на это, немногочисленные исследования посвящены проблеме прогнозирования задержек проекта на стадии планирования, что остается сложной областью. Хотя значительное количество исследований было посвящено оценке затрат на проекты, мало попыток было предпринято для изучения предсказания задержек с использованием исторических данных.

По данным исследования [4] авторами была создана диаграмма (Рис. 1), которая иллюстрирует причины задержек в строительстве в 2022 году. Причины распределены следующим образом:

- Плохое/нереалистичное планирование: 31.4%
- Проблемы с ресурсами: 19%
- Проблемы с информацией: 16%
- Изменения в спецификации: 9.9%
- Производительность ресурсов: 7.1%
- Погодные условия: 4.4%
- Финансовые задержки: 4.4%

На диаграмме причины представлены в форме круговой диаграммы, что позволяет соотнести долю каждой причины в общем количестве задержек.

## Reasons for Construction delays in '22

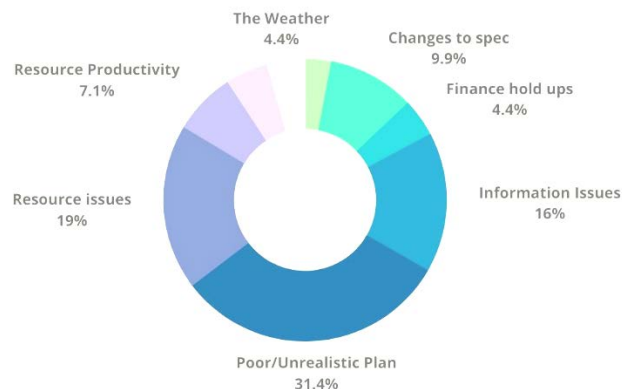


Рис. 1 – Причины задержек в строительстве

На основании представленных данных можно сделать следующие выводы:

1. Плохое/нереалистичное планирование: Этот фактор является самой значимой причиной задержек в строительстве и составляет 31.4%. Это указывает на необходимость улучшения качества планирования проектов. Проработка более реалистичных планов, с детальными временными рамками и учётом всех возможных рисков, может значительно снизить уровень задержек.
2. Проблемы с ресурсами: на втором месте по значимости находятся проблемы с ресурсами, составляющие 19%. Это может включать нехватку материалов, оборудования или рабочей силы. Улучшение управления ресурсами и создание более надёжной цепочки поставок может помочь в снижении этого фактора.
3. Проблемы с информацией: Проблемы с информацией занимают третье место с 16%. Это может включать недостаточно чёткие коммуникации между участниками проекта или ошибки в документации. Внедрение лучших систем управления информацией и повышение уровня прозрачности и коммуникации может способствовать снижению этого показателя.
4. Изменения в спецификации: Изменения в спецификации составляют 9.9%. Это указывает на необходимость более тщательной проработки проектных требований и уменьшение числа изменений в процессе реализации.
5. Производительность ресурсов: Производительность ресурсов (7.1%) также оказывает воздействие на сроки. Возможно, требуется проведение обучающих программ для увеличения квалификации рабочих и оптимизация процессов работы для повышения эффективности.
6. Погодные условия и финансовые задержки: Обе причины составляют по 4.4%. Несмотря на то, что они менее значимы по сравнению с другими факторами, стоит учитывать влияние климатических условий при планировании проектов и улучшении финансового менеджмента.

В общей сложности, эти данные подчеркивают важность комплексного подхода к управлению строительными проектами. Улучшение процессов планирования, управления ресурсами, информации и

коммуникации, а также повышение квалификации рабочих могут значительно сократить число задержек.

Несмотря на доказанный успех машинного обучения в прогнозировании затрат на строительство, мало исследований было направлено на изучение прогнозирования задержек проекта на стадии планирования. Большинство последних работ исследовали устаревшие нормативные подходы, не использующие ценные инсайты из исторических данных проектов. Таким образом, существует необходимость в разработке рекомендательной системы на основе машинного обучения, которая может предоставлять точные прогнозы задержек, оптимизируя тем самым графики проектов и процессы принятия решений.

Предсказание задержек в строительных проектах является критически важным для уменьшения перерасхода средств и обеспечения своевременного завершения проектов. Использование методов машинного обучения для анализа исторических данных предлагает перспективное решение, предоставляющее руководителям проектов прогнозные инсайты на этапе планирования. Точные прогнозы задержек могут значительно улучшить управление проектами, устраняя отклонения от графика, которые, как сообщалось, в некоторых случаях превышают 50% от плановых сроков (Gondia, A., Siam, 2020) [5]. В последние годы значительное количество исследований было сосредоточено на применении методов машинного обучения для прогнозирования различных аспектов строительных проектов, включая затраты и сроки. Например, Hashemi et al. (2020) [6] подробно изучили методы машинного обучения для оценки затрат, классифицируя их на статистические, аналоговые и аналитические модели. Аналогично, Chandanshive и Kambekar (2021) [7] разработали новые подходы с использованием моделей поддержки векторных машин (SVM) для прогнозирования затрат на строительство с высокой точностью, демонстрируя, что SVM может успешно решать широкий спектр задач в оценке строительных затрат. В последнее время увеличился интерес к исследованию методов глубокого обучения для управления строительными проектами, таких как работы Park и Yun (2023), [8] которые использовали глубокое обучение в сочетании со свойствами BIM для прогнозирования строительных затрат, подчеркивая роль машинного обучения в повышении точности прогнозов.

Saha et al. (2023) [9] в своей работе сосредоточились на исследовании факторов, влияющих на задержки в строительных проектах, особенно в контексте развивающихся экономик. Исследование использовало методы Delphi, Total Interpretive Structural Modelling (TISM) и Matrice d'Impacts Croises Multiplication Applique a Classement (MICMAC) для моделирования и анализа факторов задержек.

Исследование выделило ключевые факторы, такие как нехватка знаний о современных методах строительства и отсутствие инструментов для мониторинга проектов. Эти факторы были расположены на нижнем уровне

модели и выявлены как имеющие значительное влияние на другие задержки. Авторы подчеркивают важность применения передовых технологий и аналитических инструментов для оптимального распределения ресурсов и устранения задержек (Saha et al., 2023). [9] Предложенные методы, такие как многофакторный анализ, помогают менеджерам в принятии более обоснованных решений по управлению строительными проектами.

Второе исследование, проведенное Maуа et al. (2023) [10], направлено на разработку модели предсказания эффективности строительных проектов с использованием искусственных нейронных сетей (ANN). Основной упор был сделан на анализ и предсказание производительности проекта на основе 34 факторов, выявленных на основе мнений практиков отрасли.

Модель ANN, разработанная авторами, приняла семь входных параметров, которые были определены как наиболее влияющие на производительность. Среди этих факторов выделяются координация и приверженность участников проекта, оценка графика, опыт и доступность команды, а также поддержка высшего руководства. Модель продемонстрировала высокую точность предсказания (96.1%) и низкий уровень ошибки (3.9%), что свидетельствует о ее надежности и применимости в реальных условиях (Maуа et al., 2023) [10]

Обе статьи подчеркивают значимость применения методологий и технологий для управления и прогнозирования задержек в строительных проектах. Исследование Saha et al. (2023) [9] предлагает детализированный анализ и моделирование факторов задержек, использующих многокритериальные методы принятия решений. В то же время работа Maуа et al. (2023) [10] акцентирует внимание на использовании моделей искусственных нейронных сетей для прогнозирования эффективности проектов. Обе методики дополняют друг друга и могут быть использованы вместе для повышения точности и надежности прогнозов в строительной отрасли.

Kulkarni и соавт. (2017) в статье «Artificial neural networks for construction management: a review» [11] предоставляют обширный обзор применения ИНС в управлении строительством, отмечая их значительные преимущества в интерпретации неточной и неполной информации. Основное внимание уделяется различным типам неопределенностей, с которыми сталкивается управление строительством, включая время, стоимость, качество и безопасность. Исследователи указывают, что наиболее часто используемой архитектурой ИНС была нейронная сеть обратного распространения ошибки (feed-forward back propagation network). Однако в случае, если одной архитектуры ИНС было недостаточно, сочетание с другими инструментами машинного обучения, такими как генетическое программирование и машины опорных векторов, оказывало полезное влияние. Авторы также подчеркивают важность подлинности данных и опыта разработчика для

достижения высоких результатов (Kulkarni P. S., Londhe S. N., Deo M., 2017).

Работа Ujong и соавт. (2022) «Prediction of cost and duration of building construction using artificial neural network» [12] исследует способность искусственных нейронных сетей к предсказанию затрат и продолжительности строительных проектов. В рамках исследования были выделены шесть входных факторов (количество операций, площадь здания, тип фундамента, количество этажей, класс клиентов и подрядчиков) и два выходных параметра (продолжительность и стоимость). Результаты работы показали, что проекты, реализуемые небольшими подрядчиками, демонстрируют большую вариативность по стоимости и продолжительности из-за недостатка технических и кадровых ресурсов. Для разработки модели использовалась обратная распространённая нейронная сеть (feed-forward back-propagation network) с алгоритмом обучения Левенберга–Марквардта и критерием производительности в виде среднего квадратического отклонения (MSE). Оценка производительности моделей с использованием статистических методов показала высокую корреляцию между прогнозами ИНС и реальными данными, указывая на надёжность разработанной модели (Ujong J. A., Mbadike E. M., Alaneme G. U., 2022). [12]

Petruseva и соавт. (2016) в статье «Model for predicting construction time by using general regression neural network» [13] разработали модель предсказания времени строительства с использованием General Regression Neural Network (GRNN). Анализ данных 70 построенных зданий и опросы главных инженеров позволили построить точные прогнозы с индексом среднего абсолютного процентного отклонения (MAPE) около 2.19%, коэффициентом корреляции  $r = 0.99$  и коэффициентом детерминации  $R^2 = 0.97875$ . Разработанная модель значительно улучшает принятие решений при планировании времени строительства (Petruseva S., Pusic D. C., Pancovska V. Z., 2016). [13]

Alsugair и соавт. (2023) в статье «Artificial Neural Network Model to Predict Final Construction Contract Duration» [14] разработали ИНС модель для прогнозирования конечной длительности строительного контракта на ранних стадиях проектирования. Используя данные 135 строительных проектов в Саудовской Аравии, авторы внедрили три этапа оптимизации: стандартизацию данных, первый и второй анализы ИНС моделей. Результаты показывают высокую точность прогнозирования со средним абсолютным процентным отклонением (MAPE) 12.22%, что значительно лучше, чем у линейных регрессионных моделей. Эта работа поддерживает идею использования ИНС для принятия обоснованных решений на этапе претендерной подготовки (Alsugair A. M., et al., 2023). [14]

Yaseen и соавт. (2020) в статье «Prediction of risk delay in construction projects using a hybrid artificial intelligence model» [15] представляют гибридную модель искусственного интеллекта для прогнозирования задержек в строительных проектах. Модель сочетает

классификатор Random Forest с оптимизацией методом генетического алгоритма (RF-GA). Исследование включает анализ факторов задержек с помощью анкетного опроса и обучение модели на основе собранных данных. Модель RF-GA показала высокую точность (91.67%), kappa (87%) и низкую ошибку классификации (8.33%). Этот подход продемонстрировал свою надёжность и устойчивость в прогнозировании задержек и может быть полезен для управления строительными проектами и их устойчивости (Yaseen Z. M., et al., 2020). [15]

### III. МЕТОДОЛОГИЯ

#### A. Подготовка исходных данных

На этапе подготовки данных нашей исследовательской работы проводится тщательный разведочный анализ данных. Этот процесс включает несколько ключевых шагов:

- 1) Оценка структуры и характеристик наборов данных: Анализируем внутреннюю структуру данных, включая их типы, распределения и основные статистические характеристики.
- 2) Выявление аномалий и выбросов: Проводим детальный анализ для обнаружения и удаления аномалий и выбросов, которые могут исказить результаты.
- 3) Идентификация корреляций между переменными: Исследуем взаимосвязи между разными переменными для определения значимых корреляций, которые могут улучшить качество прогнозов.
- 4) Подготовка данных для использования методов машинного обучения: Проводим нормализацию и трансформацию данных для соответствия форматам, необходимым для использования машинного обучения.

#### B. Автоматическая разметка пользовательских наименований работ

Для однородности и улучшения качества данных, автоматическая разметка наименований работ выполняется методами машинного обучения.

Процесс включает следующие этапы

##### 1. Выбор метода:

- Кластеризация (например, K-Means, DBSCAN): обычно используются в случаях отсутствия обучающей выборки, когда необходимо группировать данные на основе их схожести. [16]

- Классификация (например, XGBoost, RandomForest, Logistic Regression): применяется при наличии ручной разметки обучающей выборки, чтобы предсказать категории новых данных. [17]

##### 2. Сравнение методов кластеризации/классификации:

- В ходе исследования проводится сравнительный анализ различных методов для выбора наилучшего подхода к разметке данных. Сравняется точность, скорость и устойчивость к аномалиям различных алгоритмов.

#### C. Построение прогнозной кривой

Для предсказания длительности работ по объектам различной площади используется построение

прогнозной кривой. Этапы включают:

1. Выбор типа кривой на графике зависимости фактической длительности работ от площади объекта строительства:

- Логарифмическая кривая: применяется при наличии экспоненциальной зависимости данных.
- Полиномиальные кривые различной степени: обычно используются для более гибкой аппроксимации зависимости.
- Линейная регрессия: применяется для базового анализа и простых линейных зависимостей.

#### *D. Выводы по разделу методологии*

Данная методология позволяет проверить гипотезу о возможности предсказания задержек в строительных проектах, опираясь на тщательно подготовленные и проанализированные данные.

#### IV. ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ

Сбор и анализ данных в строительной сфере открывают возможности для формирования рекомендаций по оптимизации длительности выполнения работ на строительных объектах. На основе исторических данных можно выявлять закономерности и тренды, которые помогут более точно прогнозировать временные затраты на выполнение различных этапов строительства. Гипотеза заключается в том, что на основе таких данных можно создавать рекомендации, которые будут способствовать более эффективному управлению строительными процессами.

##### *A. Сбор данных*

Для проверки гипотезы нами был собран обширный датасет, включающий информацию о различных строительных объектах. В общей сложности в датасет вошли данные о:

1. Размер и число объектов:
  - ~740 объектов площадью до 5 000 м<sup>2</sup>
  - ~1100 объектов площадью от 5 000 м<sup>2</sup> до 10 000 м<sup>2</sup>
  - ~430 объектов площадью более 10 000 м<sup>2</sup>
2. Стадии реализации проектов:
  - ~950 объектов находятся в стадии реализации.
  - ~1300 объектов были приняты в эксплуатацию заказчиком.
3. Объем работ:
  - В совокупности данные включают около 30 000 работ.

Такое структурирование данных позволяет учитывать специфические особенности различных типов объектов, включая их масштаб и текущую стадию строительства.

##### *B. Форма данных*

Для построения датасета были использованы следующие формы данных:

- Обезличенные данные о выполнении работ по плановым графикам: Эти данные включают временные метрики, такие как плановые, так и фактические даты начала и завершения различных строительных работ.
- Даты закрытия объемов работ по Актам выполненных работ в форме КС-2: Данные из актов выполненных работ предоставляют точечную информацию о

принятии заказчиком конкретных объемов работ, что позволяет более точно отслеживать прогресс на каждом этапе.

Состав исходных данных представлен в таблице 1.

Таблица 1 – Состав исходных данных

1	id объекта
2	Наименование объекта
3	Тип объекта
4	Плановая дата начала строительства
5	Дата завершения строительства
6	Этажность
7	Площадь
8	Сумма кол-во квартир
9	Стоимость по контракту
10	Плановая дата начала
11	Плановая дата окончания
12	Готовность на дату формирования объем
13	Готовность на дату формирования финансы
14	Фактическая дата начала
15	Фактическая дата окончания
16	Отклонение от плана на текущую дат
17	Плановая стоимость
18	Фактическая стоимость
19	Кол-во работ в графике
20	id работы
21	Наименование работы
22	Наименование классификатора вида работ
23	Единица измерения
24	Дата начала плановая
25	Дата окончания плановая
26	План физобъем
27	Признак завершения работы
28	План стоимость
29	Дата начала фактическая
30	Дата окончания фактическая
31	Процент выполнения
32	Длительность плановая
33	Фактическая длительность выполнения
34	Плановый темп работы
35	Фактический темп работы
36	Плановые трудозатраты по работе
37	Фактические трудозатраты

##### *C. Источники данных*

Источниками данных для нашего исследования служили публичные и частично обезличенные информации из государственных систем и отчетов. Основным источником данных стали:

- Публичная часть портала "ЕИС ЗАКУПКИ": Этот источник предоставляет данные по выполнению закупок и оплате работ для объектов строительства, включая жилые и социальные здания в различных субъектах Российской Федерации.

##### *D. Анализ и обработка данных*

Для обработки большого объема информации (порядка 30 000 работ) были использованы современные методы анализа данных, включая автоматизированные алгоритмы машинного обучения для классификации и предсказания временных затрат. Основные этапы анализа включали:

1. Очистка данных: Исключение дубликатов, проверка корректности заполнений и устранение явных ошибок в датах и объемах.

2. Нормализация данных: Приведение данных к единому формату, чтобы обеспечить их сопоставимость.

3. Анализ и визуализация: Использование статистических методов и инструментов визуализации данных для выявления трендов и закономерностей. Основной акцент делался на идентификацию проектов, выходящих за рамки среднестатистических показателей. На основании полученных данных можно сделать следующие выводы:

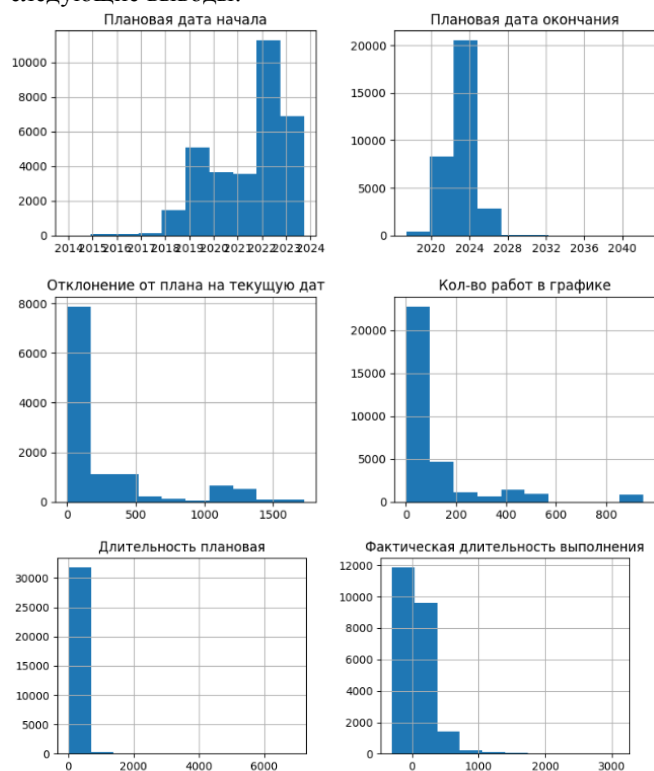


Рис. 2 – А. Плановая дата начала, Б. Плановая дата окончания, В. Отклонение от плана на текущую дату, Г. Кол-во работ в графике, Д. Длительность плановая, Е. Фактическая длительность выполнения.

1. Плановая дата начала (Рис. 2 – А):

- Пик планируемых начал проектов наблюдается в периоды с 2018 по 2024 годы.

- Наибольший всплеск начала проектов отмечается в последние годы (2022-2024).

2. Плановая дата окончания (Рис. 2 – Б):

- Пик планируемых окончаний проектов приходится на период 2020-2024 годы.

- Большинство проектов планировалось закончить в эти годы, что указывает на высокую активность в планировании проектов на недавно прошедший период.

3. Отклонение от плана на текущую дату (Рис. 2 – В):

- Большинство проектов имеют небольшие отклонения от плана на текущую дату.

- Значительное количество проектов имеют отклонения до 500 дней, что указывает на существующие проблемы с соблюдением сроков.

4. Количество работ по графику Рис. 2 – Г):

- Большинство проектов имеют небольшое количество запланированных работ.

- Большинство планов содержат до 200 задач, в то время как некоторые планы имеют значительно больше задач, что может указывать на разную степень сложности проектов.

5. Длительность плановая (Рис. 2 – Д):

- Плановая длительность большинства проектов находится в диапазоне до 500 дней.

- Пики плановой длительности наблюдаются до 1000 дней, однако большинство проектов планируется к выполнению в существенно меньшие сроки.

6. Фактическая длительность выполнения Рис. 2 – Е):

- Большинство проектов фактически завершается в сроки до 500 дней.

- Наблюдаются случаи, когда фактическая длительность выполнения проектов превышает плановую, указывая на потенциал для оптимизации процессов.

#### Е. Выводы по количественным показателям

- Планирование и выполнение: на графиках наблюдается значительная концентрация плановых начал и окончаний проектов в период 2018-2024 годов. Это указывает на высокий темп проектной деятельности в эти годы.

- Отклонения от плана: Большинство проектов имеют отклонения от плана, что требует анализа и корректировки процессов планирования и управления проектами для повышения точности прогноза.

- Длительность проектов: Фактическая длительность выполнения проектов часто превышает плановую, указывая на возможное недооценивание временных затрат или недостаточное управление ресурсами и рисками.

Резюмируя, существует потенциал для улучшения процесса планирования, управления ресурсами и контроля за выполнением проектов, что позволит уменьшить отклонения от плана и более точно соблюдать сроки.

#### Ф. Анализ по категориям объектов и составу данных

На Рис. 3 показаны различные классы объектов и стадии их выполнения.

Класс объекта ML	Стадия ML	
Автомобильные дороги и искусственные сооружения	Благоустройство	4
	Капитальный ремонт	20
	Реконструкция	66
	Ремонт	19
	Строительство	33
Газификация	Строительство	72
ЖКХ	Благоустройство	174
	Капитальный ремонт	251
	Реконструкция	3
	Ремонт	1
Жилищное строительство	Строительство	6
	Реконструкция	3
Капитальный ремонт	Строительство	195
	Капитальный ремонт	3
Объекты благоустройства	Благоустройство	14
	Реконструкция	3
	Строительство	7
Объекты здравоохранения	Строительство	17
Объекты культуры	Реконструкция	7
	Строительство	26
Объекты образования	Реконструкция	1
	Строительство	27
Объекты связи	Строительство	2
	Капитальный ремонт	1
Сооружения водоснабжения	Реконструкция	26
	Строительство	38
	Капитальный ремонт	3
Сооружения теплоснабжения	Реконструкция	1
	Строительство	3
	Капитальный ремонт	3

Рис. 3 – Категории объектов из состава данных

Из таблицы (Рис. 3) можно сделать несколько выводов по анализу набора данных в контексте категорий объектов:

#### 1. Высокая активность в ЖКХ:

- Большинство работ в классе "ЖКХ" сосредоточены на благоустройстве (174 объекта) и капитальном ремонте (251 объект). Это указывает на повышенное внимание к улучшению и поддержке инфраструктуры ЖКХ.

#### 2. Жилищное строительство:

- В категории "Жилищное строительство" основная часть работ приходится на строительство (195 объектов) с небольшим участием реконструкции (3 объекта). Это свидетельствует о значительном количестве новых жилищных проектов.

#### 3. Развитие объектов здравоохранения и культуры:

- В области здравоохранения проводится 17 проектов по строительству.

- В области культуры основная активность сосредоточена также на строительстве (26 объектов) и реконструкции (7 объектов).

#### 4. Инфраструктурные проекты:

- В области "Газификация" активно ведется строительство (72 объекта).

- Объекты водоснабжения и теплоснабжения также имеют некоторые активности в категориях капитального ремонта, реконструкции и строительства.

#### 5. Разнообразие стадий осуществления проектов:

- Стадии выполнения работ варьируются от благоустройства и капитального ремонта до реконструкции и строительства. В автомобильных

дорогах и искусственных сооружениях, помимо строительства (33 проекта), активно происходят реконструкция (66 проекта) и капитальный ремонт (20 проектов).

#### 6. Общие проекты благоустройства и капитального ремонта:

- На проекты по благоустройству и капитальному ремонту также отводится значительное количество ресурсов во всех категориях: объекты благоустройства (14 благоустройств), капитальный ремонт (3 капитальных ремонта).

#### G. Выводы по категориям объектов и составу данных

- Акцент на улучшение инфраструктуры: Значительные усилия и ресурсы направлены на улучшение инфраструктуры ЖКХ и строительство новых жилых объектов.

- Объекты социальной направленности: Здравоохранения и культурные объекты также являются приоритетами, с фокусом на их капитальное улучшение и расширение.

- Потребность в новых построенных и реконструированных объектах: Разнообразие проектов свидетельствует о системном подходе к обновлению и развитию объектов инфраструктуры, включая дороги, газификацию, водоснабжение и теплоснабжение.

#### H. Промежуточные результаты и выводы

Первичный анализ данных показал наличие значительной вариативности в длительности выполнения работ на различных этапах строительства. Результаты из тестов на нормальность распределений подтвердили, что данные не являются нормальными, что свойственно для реальных строительных проектов, где множество факторов может повлиять на временные затраты.

На основе этих данных была разработана модель, использующая методы машинного обучения для прогнозирования длительности выполнения различных строительных этапов. В результате мы смогли сформулировать рекомендации, которые помогают оптимизировать планирование и управление строительными проектами, учитывая специфику конкретных объектов и этапов.

#### V. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА И ФИЛЬТРАЦИЯ ИСХОДНЫХ ДАННЫХ

Для разработки моделей прогнозирования и оптимизации длительности выполнения работ необходимо было провести детальный анализ собранного массива данных о проектном исполнении. Фокус исследования был направлен на выявление и устранение аномалий, изучение закономерностей в распределении сроков выполнения работ и автоматизацию их классификации. В результате анализа подтвердились возможности прогнозирования плановых сроков на основе фактических данных.

#### A. Обработка и фильтрация данных

В процессе анализа были выявлены работы с явно аномальными временными характеристиками, которые потенциально могли исказить результаты:

- Удалены работы, фактический срок выполнения которых:

- Превышал плановый срок на 300% и более. Такие случаи могут быть результатом чрезвычайных обстоятельств, неверно введенных данных или других разовых происшествий, что делает их нерепрезентативными для общей модели.
- Выполнялся на 50% быстрее планового срока. Работа, завершенная значительно быстрее планируемого времени, также может свидетельствовать о неправильном планировании или ошибках в данных.

Этот этап очистки данных позволил получить более однородную выборку для дальнейшего анализа.

**V. Подтверждение гипотезы**

После фильтрации была проведена оценка распределения длительностей завершенных работ относительно их плановых сроков. Основные результаты анализа включают:

- Более 50% работ в среднем отстают в 2-2.5 раза от плановых сроков. Это указывает на систематические проблемы с планированием сроков в строительных проектах. Данный вывод подтолкнул к дальнейшему исследованию факторов, влияющих на такие задержки, и разработке методов их прогнозирования и предотвращения.

На Рис. 4 отображены две категории объектов, в которых продемонстрировано отставание.

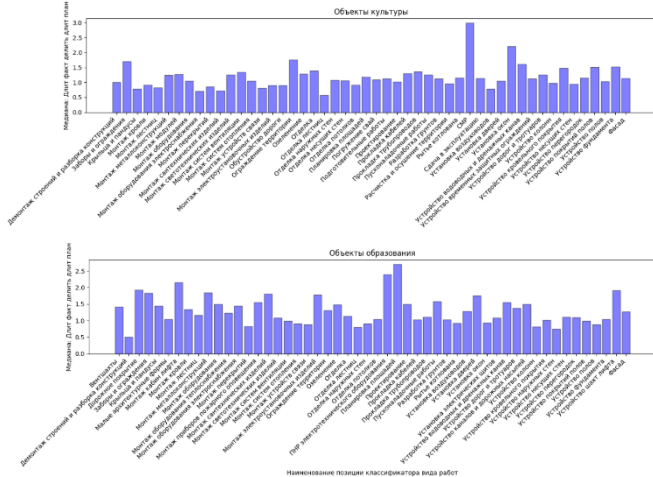


Рис. 4 – А. Объекты культуры, Б. Объекты образования. Выявленные закономерности будут интегрированы в разработку решения для руководителей проектов.

Основные функциональные возможности такого решения включают:

- Прогнозирование срывов сроков: Благодаря модели прогнозирования на основе исторических данных, руководители смогут заранее узнать о возможных задержках.
- Оптимизация планирования: Система будет предоставлять рекомендации по оптимизации планов, что может существенно уменьшить количество случаев задержек.
- Автоматическая классификация: Обеспечение автоматической разметки данных позволит повысить точность анализа и улучшить качество прогнозов.

**VI. АНАЛИЗ РАСПРЕДЕЛЕНИЙ ФАКТИЧЕСКИХ ДЛИТЕЛЬНОСТЕЙ РАБОТ**

При анализе большого объема данных в строительной сфере, особенно касающихся фактической длительности выполнения определенных строительных задач, критически важно понимать природу распределения этих данных. Констатируем факт, что у нас имеется 30 тысяч работ, по каждой из которых отражено распределение длительности выполнения в различных проектах или контекстах (например, газификация, ЖКХ, жилищное строительство).

Построив по каждой из работ графики (Рис. 4) распределения можно сделать следующие выводы:

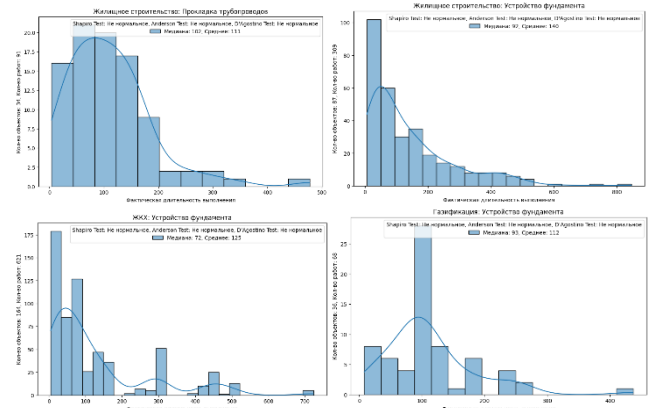


Рис. 5 – Графики распределения по четырем работам

1. Неравномерное распределение: Во всех случаях мы наблюдаем ярко выраженное неравномерное распределение данных, характеризующееся левосторонней асимметрией. Это свидетельствует о том, что большинство проектов завершается в относительно короткие сроки, в то время как небольшая часть проектов имеет значительно большую продолжительность исполнения.

2. Отклонение от нормальности: Статистические тесты (Shapiro, Anderson и D'Agostino) подтверждают, что распределения данных не являются нормальными. Это согласуется с наблюдаемыми правыми "хвостами" на гистограммах и отклонениями от красной линии на Q-Q графике (Рис. 5)

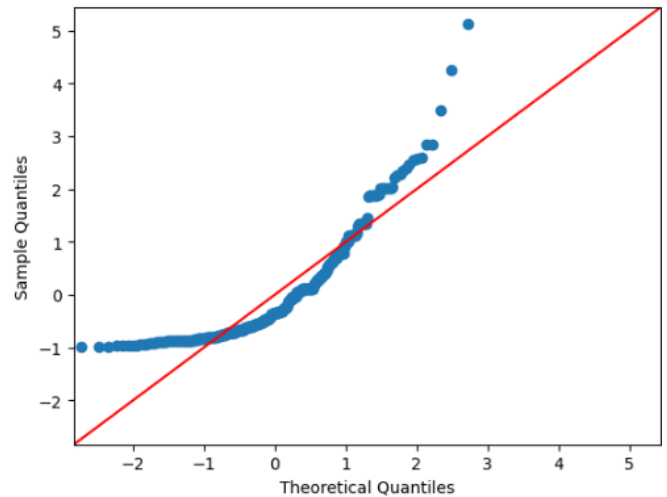


Рис. 5 - График квантиль-квантиль (Q-Q)

3. Асимметрия: Такие показатели, как медиана и среднее значение, еще раз подтверждают асимметрию.



Средние значения во всех случаях значительно превышают медианы, что характерно для распределений, имеющих длинные правые хвосты.

Осознание отсутствия нормального распределения имеет несколько важных последствий: 1. Выбор методов анализа: Многие статистические методы и инструменты, такие как методы линейной регрессии, предполагают нормальное распределение данных. Использование таких методов на данных, отклоняющихся от нормального распределения, может привести к ошибочным выводам. В данном случае нужно использовать иные методы анализа, такие как методы, подходящие для работы с несимметричными данными (например, логнормальное распределение или методики преобразования данных).

Понимание распределения длительности выполнения задач может помочь в лучшем управлении проектами. Асимметрия данных говорит о том, что необходимо закладывать значительные временные резервы для части проектов, которые могут задерживаться.

Для улучшения точности прогнозирования длительности выполнения будущих проектов и оценки рисков необходимо уделять внимание тем проектам, которые выходят за пределы основного распределения. Это может включать дополнительное ресурсное обеспечение и усиление контроля за такими проектами. В условиях работы с большими объемами данных критически важно проводить детализированный анализ распределений длительности выполнения задач.

Отсутствие нормального распределения требует использования специализированных методов анализа и прогнозирования, которые учитывают асимметрию данных и повышенную вероятностную плотность на одном из концов. Такие методы позволят улучшить точность прогнозов, управление проектами и оценку рисков, что, в конечном итоге, приведет к более успешному выполнению строительных проектов.

## VII. АВТОМАТИЧЕСКАЯ РАЗМЕТКА МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ ПОЛЬЗОВАТЕЛЬСКИХ НАИМЕНОВАНИЙ РАБОТ ДЛЯ ОДНОРОДНОСТИ И ПРИВЕДЕНИЯ К ЕДИНОМУ КЛАССИФИКАТОРУ РАБОТ

### A. Неоднородность наименований работ

Разведочный анализ данных, в частности анализ графиков нормального распределения работ, позволил выявить неоднородность в наименованиях строительных работ, что потребовало применения автоматической разметки. Это необходимо для:

- Приведения наименований строительных работ к единому классификатору, что способствует стандартизации и улучшению качества данных.
- Использования средств машинного обучения для автоматической классификации пользовательских наименований строительных работ. Разработанные модели машинного обучения позволяют свести разнородные данные к единому стандарту, облегчая анализ и прогнозирование.

### B. Применение методов машинного обучения для автоматической переработки наименований работ

Эффективное управление большими объемами данных в строительных проектах требует стандартизации и унификации наименований работ. В рамках нашего исследования была разработана стратегия автоматической разметки наименований строительных работ, которая задействует методы машинного обучения. В результате был сформирован единый классификатор, включающий 106 позиций, на основе и сформирована вручную размеченная обучающая выборка на 4000 наименований работ, на основе которой было размечено 32 000 наименований работ. В данном разделе описан процесс выбора методов, формирования обучающей выборки и сравнительный анализ различных алгоритмов машинного обучения.

### C. Формирование единого классификатора

Для стандартизации данных и обеспечения единой структуры наименований строительных работ был создан и утвержден классификатор, состоящий из 106 позиций. Этот классификатор служит основой для автоматической и точной разметки всех поступающих данных, исключая вариативность в названиях, которая могла бы затруднить дальнейший анализ и прогнозирование. После согласования классификатора была сформирована обучающая выборка на 4000 позиций.

### D. Разметка наименований

На основе разработанного классификатора и обучающей выборке была выполнена автоматическая разметка 32 000 наименований строительных работ. Этот процесс позволил значительно улучшить качество данных и упростить обработку для дальнейших этапов анализа и прогнозирования.

### E. Выбор метода машинного обучения

При выборе методов машинного обучения для автоматической разметки наименований строительных работ рассматривались два основных подхода: кластеризация и классификация.

### F. Методы кластеризации

Методы кластеризации, такие как K-Means и DBSCAN [16], были первоначально рассмотрены для автоматической разметки данных. Преимущества и проблемы данного подхода включают:

- Отсутствие обучающей выборки: методы кластеризации не требуют заранее размеченных данных, что является существенным плюсом, особенно при работе с большими объемами неструктурированной информации.
- Сложности интерпретации: несмотря на автоматическую группировку данных, результаты кластеризации часто требуют дополнительной интерпретации и валидации экспертов.

### G. Методы классификации

Для методов классификации, таких как XGBoost, RandomForest и Logistic Regression [17], была сформирована обучающая выборка с помощью ручной



преимущества и ограничения. Дополнение пропущенных значений на графиках (Рис. 7) позволило нам создать непрерывные модели, способные предоставить более точные прогнозы длительности работ для объектов различной площади.



Рис. 7 – Пример прогнозной кривой

Эти результаты могут быть использованы для оптимизации планирования и управления строительными проектами на основании машинного обучения, позволяя более точно оценивать сроки выполнения работ в зависимости от площади объекта или других факторов, о которых будет рассказано в дальнейшем.

## IX. ЗАКЛЮЧЕНИЕ

В качестве источника исходных данных были использованы обезличенные фактические данные о выполнении работ по плановым графикам, даты закрытия объемов работ по Актам выполненных работ форма КС-2, записи в общих журналах работ форма КС-6 и других бумажных документах. Источником выступали данных из открытых государственных систем по выполнению нескольких десятков тысяч работ нескольких сотен объектов строительства, а именно по жилым и социальным зданиям.

В качестве методологии исследования был выполнен разведочный анализ данных, позволяющий оценить структуру и характеристики наборов данных, выявить аномалии и выбросы, идентифицировать корреляции между переменными и подготовить данные для использования методов машинного обучения. В качестве инструментов были использованы гистограммы, тепловые карты, статистические метрики (например, среднее отклонение и медиана) и корреляционный анализ.

В результате в собранном массиве данных о проектном исполнении были выделены работы с повторяющимися закономерностями. Удалены работы, фактический срок выполнения которых более, чем в 3 раза больше и на 50% быстрее планового срока.

Выполнен анализ распределений длительностей завершённых работ относительно плановых сроков. Для продолжения дальнейшего прогнозирования срывов сроков строительства была выявлена необходимость в выполнении автоматической разметки (классификации) пользовательских наименований строительных работ средствами машинного обучения для приведения их к единому классификатору.

Выявленные закономерности будут использованы для разработки решения, позволяющего осуществлять прогнозирование срывов сроков на реальных объектах строительства с целью предоставления руководителям информации о прогнозных сроках выполнения работ на

этапе планирования.

Апробация будет произведена в результате внедрения информационной системы аналитики сроков и статусов объекта строительства.

## X. ПЕРСПЕКТИВЫ РАЗВИТИЯ

Разработка прогнозных моделей для зависимости фактической длительности строительных работ от различных параметров является важным шагом в оптимизации процессов строительства. Однако на этом возможности нашей системы не исчерпываются. В данной секции будут рассмотрены направления дальнейшего развития, направленные на создание более эффективных и комплексных инструментов для управления строительными проектами.

### A. Рекомендательная система

Одним из перспективных направлений является разработка интерфейса рекомендательной системы. Это будет интерактивное приложение, позволяющее пользователям вводить наименования конкретных работ и их плановую длительность, на выходе предоставляя рекомендации по фактическим срокам завершения работы:

- Ввод информации: Пользователь вводит наименование работы и плановую длительность.
- Обработка данных: Система автоматически определяет классификатор работы на основе введенного наименования.
- Результат: Пользователь получает рекомендованные фактические сроки завершения работы, опираясь на исторические данные и прогнозные модели.

### B. Дополнительные атрибуты

Для повышения точности прогнозов и релевантности рекомендаций, можно использовать дополнительные группирующие атрибуты помимо площади объекта:

- Стоимость объекта: Анализ зависимости длительности работ от общей стоимости строительства.
- Стоимость работ: Учет затрат на выполнение конкретных работ, что может напрямую влиять на их длительность.
- Погодные условия: Включение данных о погодных условиях, которые могут существенно влиять на сроки выполнения работ.
- Критический путь: Определение работ, которые лежат на критическом пути проекта, и учет их влияния на общую длительность.
- Время года: Анализ сезонных факторов и их влияние на сроки выполнения строительных работ.
- Прочие строительные признаки: Включение других значимых параметров, таких как доступность материалов, квалификация рабочей силы и техники оснащение.

### C. Автоматизация и интеграция

В результате разработки данных механизмов мы стремимся создать систему, которая будет обладать следующими функциональными возможностями:

- Определение классификатора: Автоматическое присвоение классификатора работы на основании ее

наименования.

- Исторический анализ: Доступ к историческим данным о длительности выполнения аналогичных работ.
- Интерактивный интерфейс: Интуитивный интерфейс для ввода исходных данных и получения рекомендаций.
- Многомерный анализ: Возможность учета множества параметров для более точных прогнозов и рекомендаций.
- Машинное обучение: Использование алгоритмов машинного обучения для постоянного улучшения модели и адаптации к новым данным.

#### D. Выводы

Разработка и внедрение данных инструментов и методов позволит значительно повысить эффективность планирования и управления строительными проектами. Это в свою очередь обеспечит более точные прогнозы, позволив минимизировать риски и оптимизировать использование ресурсов. Создание рекомендательной системы и интеграция дополнительных группирующих атрибутов сделает процессы более управляемыми и предсказуемыми, что приведет к успешному завершению проектов в установленные сроки и в рамках бюджета.

#### БИБЛИОГРАФИЯ

- [1] Распоряжение Правительства РФ от 31 октября 2022 г. № 3268-р
- [2] Давыдова К. А. Оценка резерва времени, необходимого для предотвращения срывов сроков строительного производства //Техника. Технологии. Инженерия. – 2018. – №. 3. – С. 29-31.
- [3] Харисов А. Р., Коклюгина Л. А., Коклюгин А. В. Исследование существующих методов определения продолжительности строительства промышленных объектов // Известия КазГАСУ. 2012. №1 (19)
- [4] Jim Burns DELAYS IN THE CONSTRUCTION INDUSTRY: OUR 2022 SURVEY RESULTS AND HOW THEY COMPARE TO 2016 [Электронный ресурс <https://www.cornerstoneprojects.co.uk/blog/delays-in-the-construction-industry-our-2022-survey-results-and-how-they-compare-to-2016/>, дата обращения 25.06.2024)
- [5] Gondia, A., Siam, A., El-Dakhkhni, W., & Nassar, A. H. (2020). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management*, 146(1).
- [6] Tayefeh Hashemi, S., Ebadati, O. M., & Kaur, H. (2020). Cost estimation and prediction in construction projects: A systematic review on machine learning techniques. *SN Applied Sciences*, 2(10), 1703.
- [7] Chandanshive, V. B., & Kambekar, A. R. (2021). Prediction of building construction project cost using support vector machine. *Industrial Engineering and Strategic Management*, 1(1), 31-42.
- [8] Park, D., & Yun, S. (2023). Construction Cost Prediction Using Deep Learning with BIM Properties in the Schematic Design Phase. *Applied Sciences*, 13(12), 7207.
- [9] Saha, Saibal Kumar & Patil, Anchal & Dwivedi, Ashish & Pamucar, Dragan & Pillai, Aparna. (2023). Analyzing the interactions among delay factors in construction projects: A multi criteria decision analysis. *Reports in Mechanical Engineering*. 4. 241-255. 10.31181/rme040116112023s.
- [10] Maya, R., Hassan, B., & Hassan, A. (2023). Develop an artificial neural network (ANN) model to predict construction projects performance in Syria. *Journal of King Saud University-Engineering Sciences*, 35(6), 366-371.
- [11] Kulkarni P. S., Londhe S. N., Deo M. Artificial neural networks for construction management: a review //Journal of Soft Computing in Civil Engineering. – 2017. – Т. 1. – №. 2. – С. 70-88.
- [12] Ujong J. A., Mbadike E. M., Alaneme G. U. Prediction of cost and duration of building construction using artificial neural network //Asian Journal of Civil Engineering. – 2022. – Т. 23. – №. 7. – С. 1117-1139.
- [13] Petruseva S., Pusic D. C., Pancovska V. Z. Model for predicting construction time by using general regression neural network //International Scientific Conference People, Buildings and Environment. – 2016. – Т. 29. – С. 33-46.
- [14] Alsugair A. M. et al. Artificial Neural Network Model to Predict Final Construction Contract Duration //Applied Sciences. – 2023. – Т. 13. – №. 14. – С. 8078.
- [15] Yaseen Z. M. et al. Prediction of risk delay in construction projects using a hybrid artificial intelligence model //Sustainability. – 2020. – Т. 12. – №. 4. – С. 1514.
- [16] Гура Д.А., Болтовнина О.С. ПРИМЕНЕНИЕ СУЩЕСТВУЮЩИХ АЛГОРИТМОВ ПО КЛАССИФИКАЦИИ И КЛАСТЕРИЗАЦИИ ТОЧЕК ЛАЗЕРНОГО ОТРАЖЕНИЯ (K-MEANS, DBSCAN, SVM) ДЛЯ РЕШЕНИЯ ЗАДАЧ ТЕРРИТОРИАЛЬНОГО ПЛАНИРОВАНИЯ // Вестник ДГТУ. Технические науки. 2023. №1. URL: <https://cyberleninka.ru/article/n/primeneniye-suschestvuyuschih-algoritmov-po-klassifikatsii-i-klasterezatsii-tochek-lazernogo-otrazheniya-k-means-dbscan-svm-dlya-data-obrasheniya> (дата обращения: 30.06.2024).
- [17] Низамитдинов А. И., Джаборова Ш. А. СОВРЕМЕННЫЕ МЕТОДЫ КЛАССИФИКАЦИИ ТЕКСТА НА ОСНОВЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ //ВАЗОРАТИ МАОРИФ ВА ИЛМИ ЧУМХУРИИ ТОЧИКИСТОН Донишкадаи политехникии Донишгоҳи техникии Тоҷикистон ба номи академик МС Осимӣ МАҶАЛЛАИ ИЛМӢ-ТЕХНИКӢ «ПАЁМИ ДПДТТ. – 2022. – С. 23.

# Predicting construction delays using machine learning based on historical data on the actual duration of completed projects

V.V. Konkov, V.I. Shirokov, M.G. Zhabitsky

**Abstract** — This study examines the problem of exceeding planned construction deadlines, which is a significant obstacle to increasing the efficiency of the construction industry and its contribution to the economy of the Russian Federation. As part of the task defined by the development strategy of the construction industry and housing and communal services of Russia for the period until 2030, it is envisaged to reduce the duration of the investment and construction cycle by 30%. However, current planning methods based on outdated regulatory approaches have shown to be ineffective due to inattention to the statistical data of already completed projects.

To solve this problem, the authors proposed a hypothesis about creating a system of recommendations based on the analysis of historical data on the implementation of individual construction works and construction projects. The initial data was anonymized information about the planned and actual deadlines for completing work, processed by methods of exploratory data analysis and machine learning. The results obtained made it possible to identify patterns in the implementation of construction projects and develop a methodology for predicting delays in individual works, aimed at optimizing planning and reducing the duration of construction projects planned for implementation.

The implementation of the proposed approach and its testing on real construction projects provide managers with timely information to adjust schedules and improve project management efficiency.

**Keywords** - Construction delays, machine learning, predictive modeling, historical data analysis, project management.

## REFERENCES

- [1] Order of the Government of the Russian Federation of October 31, 2022 No. 3268-r
- [2] Davydova K. A. Estimation of the time reserve required to prevent delays in construction production // *Technology. Technologies. Engineering.* – 2018. – No. 3. – pp. 29-31.
- [3] Kharisov A.R., Koklyugina L.A., Koklyugin A.V. Study of existing methods for determining the duration of construction of industrial facilities // *News of KazGASU.* 2012. No. 1 (19)
- [4] Jim Burns DELAYS IN THE CONSTRUCTION INDUSTRY: OUR 2022 SURVEY RESULTS AND HOW THEY COMPARE TO 2016 [Electronic resource <https://www.cornerstoneprojects.co.uk/blog/delays-in-the-construction-industry-our-2022-survey-results-and-how-they-compare-to-2016/>, accessed 06/25/2024)
- [5] Gondia, A., Siam, A., El-Dakhkhni, W., & Nassar, A. H. (2020). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management*, 146(1).
- [6] Tayefeh Hashemi, S., Ebadati, O. M., & Kaur, H. (2020). Cost estimation and prediction in construction projects: A systematic review on machine learning techniques. *SN Applied Sciences*, 2(10), 1703.
- [7] Chandanshive, V. B., & Kambekar, A. R. (2021). Prediction of building construction project cost using support vector machine. *Industrial Engineering and Strategic Management*, 1(1), 31-42.
- [8] Park, D., & Yun, S. (2023). Construction Cost Prediction Using Deep Learning with BIM Properties in the Schematic Design Phase. *Applied Sciences*, 13(12), 7207.
- [9] Saha, Saibal Kumar & Patil, Anchal & Dwivedi, Ashish & Pamucar, Dragan & Pillai, Aparna. (2023). Analyzing the interactions among delay factors in construction projects: A multi criteria decision analysis. *Reports in Mechanical Engineering.* 4. 241-255. 10.31181/rme040116112023s.
- [10] Maya, R., Hassan, B., & Hassan, A. (2023). Develop an artificial neural network (ANN) model to predict construction projects performance in Syria. *Journal of King Saud University-Engineering Sciences*, 35(6), 366-371.
- [11] Kulkarni P. S., Londhe S. N., Deo M. Artificial neural networks for construction management: a review // *Journal of Soft Computing in Civil Engineering.* – 2017. – T. 1. – No. 2. – pp. 70-88.
- [12] Ujong J. A., Mbadike E. M., Alaneme G. U. Prediction of cost and duration of building construction using artificial neural network // *Asian Journal of Civil Engineering.* – 2022. – T. 23. – No. 7. – pp. 1117-1139.
- [13] Petrusseva S., Pusic D. C., Pancovska V. Z. Model for predicting construction time by using general regression neural network // *International Scientific Conference People, Buildings and Environment.* – 2016. – T. 29. – P. 33-46.
- [14] Alsugair A. M. et al. Artificial Neural Network Model to Predict Final Construction Contract Duration // *Applied Sciences.* – 2023. – T. 13. – No. 14. – P. 8078.
- [15] Yaseen Z. M. et al. Prediction of risk delay in construction projects using a hybrid artificial intelligence model // *Sustainability.* – 2020. – T. 12. – No. 4. – P. 1514.
- [16] Gura D.A., Boltovnina O.S. APPLICATION OF EXISTING ALGORITHMS FOR CLASSIFICATION AND CLUSTERIZATION OF LASER REFLECTION POINTS (K-MEANS, DBSCAN, SVM) TO SOLVING TERRITORIAL PLANNING PROBLEMS // *Vestnik DSTU.* Technical science. 2023. No. 1. URL: <https://cyberleninka.ru/article/n/primeneniye-suschestvuyuschih-algoritmov-po-klassifikatsii-i-klasterizatsii-tochek-lazernogo-otrazheniya-k-means-dbscan-svm-dlya> (access date: 06/30/2024).
- [17] Nizamitdinov A. I., Jaborova Sh. A. MODERN METHODS OF TEXT CLASSIFICATION BASED ON MACHINE LEARNING ALGORITHMS // *VAZORATI MAORIF VA ILMI GUMURIY TOKIKISTON* Donishkadai Polytechnic Donishgoi Technology Tokikiston ba nomi akademician MS Os imʼi MAʼQALLAI ILMII-TEKNIK “PAYOMI DPDTT. – 2022. – P. 23.