

Математические модели и программный комплекс для анализа контрагентов в системе прогнозирования исполнения государственных контрактов

Д.Ю. Рубцов

Финансовый университет при Правительстве Российской Федерации, Москва

Аннотация: Разработаны модели и программный комплекс, позволяющие производить анализ контрагентов на предмет вероятности исполнения государственных контрактов. Проведен сравнительный анализ моделей машинного обучения: логистической регрессии, леса решений, кластеризации и нейросети. Разработан программный комплекс, позволяющий проводить прогнозирование контрактов. Проведен вычислительный эксперимент анализ контрагентов с учетом исполненных, либо не завершенных ими контрактов. Установлена лучшая модель, демонстрирующая точность прогноза 97,89% по метрике точность.

Ключевые слова: математическое моделирование, программный комплекс, интеллектуальные модели, финансовый сектор, государственные контракты, информационная инфраструктура.

Введение

Анализ исполнения государственных контрактов является сложной и многоуровневой задачей, качество исполнения которой влияет на общие потери государства от данной области. В связи с высоким уровнем развития интеллектуальных систем и машинного обучения, такие аналитические задачи необходимо передавать для выполнения компьютеру, оставляя человека только для контроля [1-3]. Системы, построенные на машинном обучении в таких областях в разы эффективнее человека, но их моделирование и имплементация отдельная непростая задача. Для её упрощения, необходимо разбиение на различные модули. Один из таких модулей должен заниматься анализом контрагентов-исполнителей государственных контрактов.

Модуль анализа контрагентов является важной частью системы прогнозирования исполнения государственных контрактов. Он позволит выявить влияние характеристик контрагентов на вероятность исполнения того или иного контракта. Такой подход с разбиением системы на модули не

только эффективен в рамках моделирования, но и облегчает задачу с разработкой и восприятием самой системы человеком.

В данной работе предлагается несколько подходов к разработке системы прогнозирования исполнения государственных контрактов, а также представлена методология разработки модуля анализа контрагентов для данной системы, позволяющая заранее выявить положительное или негативное влияние исполнителя на тот или иной контракт. Основные задачи модуля включают оценку финансовой устойчивости контрагентов, проверку их репутации и анализ исторических данных. Оценка финансовых показателей, таких, как выручка и ликвидность, помогает определить способность контрагента выполнять свои обязательства. Важно также собрать информацию о предыдущих контрактах и наличии штрафов или судебных разбирательств, что позволит выявить потенциальные риски.

Модуль анализа контрагентов должен быть интегрирован с внешними базами данных, такими как реестры юридических лиц и данные Федеральной налоговой службы. Это обеспечит доступ к актуальной информации [4]. Автоматизированный анализ, с использованием методов машинного обучения, таких как классификация и регрессия, позволяет обрабатывать большие объемы данных и прогнозировать риски на основе паттернов, выявленных в прошлом. Например, алгоритмы могут анализировать историческую статистику выполнения контрактов и определять вероятность успешного завершения новых сделок. Кроме того, важным аспектом является визуализация данных. Генерация наглядных отчетов и дашбордов поможет быстро интерпретировать результаты анализа и облегчить принятие решений. Оценка рисков станет более обоснованной благодаря разработке системы, учитывающей различные показатели, такие как вероятность неплатежа и финансовые проблемы.

Внедрение модуля анализа контрагентов дает множество преимуществ. Это не только снижает риски, связанные с заключением договоров с ненадежными исполнителями, но и оптимизирует процесс закупок [5,6]. Быстрый доступ к информации о контрагентах и их оценка позволяют оперативно находить надежных поставщиков, что значительно ускоряет процесс принятия решений. Кроме того, открытость информации способствует формированию честной конкурентной среды.

Методология разработки интеллектуальной системы

Предложенная интеллектуальная система базируется на обучении с использованием открытых ретроспективных данных о контрактах, содержащих информацию о закупках, проводимых ведомствами и государственными корпорациями Российской Федерации в соответствии с Федеральным Законом «О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд» от 05.04.2013 N 44-ФЗ [7]. Система также учитывает данные о проведенных транзакциях исполнителей данных контрактов и опирается на набор критериев, которые позволяют оценить вероятность успешного исполнения каждого отдельного контракта [8,9]. Схема, иллюстрирующая полную систему прогнозирования исполнения государственных контрактов, представлена на рисунке 1.

Разработка модуля анализа контрагентов требует системного подхода, основанного на сочетании аналитических, технологических и организационных методов. Основная цель — создать интеллектуальную систему, которая будет эффективно обрабатывать и анализировать данные, обеспечивая надежные прогнозы исполнения государственных контрактов.

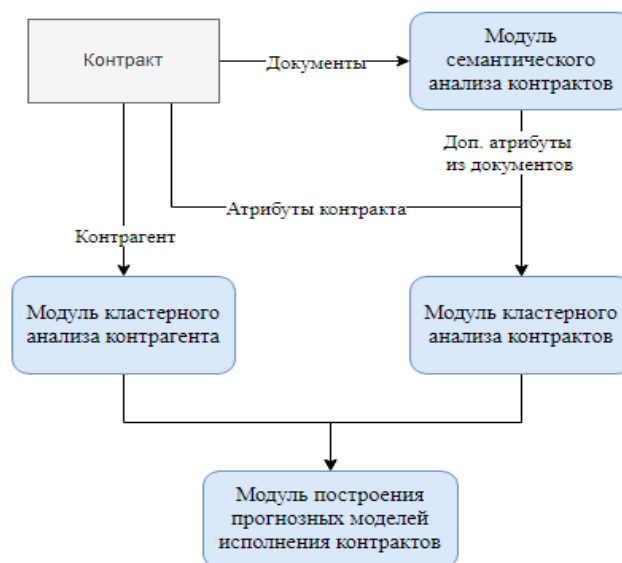


Рис. 1. – Схема программного комплекса прогнозирования исполнения контрактов.

1. Определение требований. На начальном этапе важно четко определить требования к системе. Это включает в себя формулирование задач, которые она должна решать, и функций, которые должны быть реализованы. Консультации с экспертами в области закупок, юриспруденции и анализа данных помогут понять, какие аспекты необходимо учитывать, чтобы система была максимально полезной. Здесь можно попытаться определить какую задачу будет выполнять данный модуль. Его можно использовать как полностью отдельную структуру, результаты которого можно использовать как отдельные параметры для будущего принятия решения в итоговых моделях прогнозирования исполнения государственных контрактов. Так же можно принять решения о том, чтобы сами параметры были использованы непосредственно сразу в финальном модуле.

2. Сбор данных. Необходимо создать обширную базу информации о контрагентах, включающую финансовые показатели, репутационные данные, историю исполнения контрактов и другую релевантную информацию. Важно обеспечить интеграцию с внешними источниками данных, такими как

реестры юридических лиц и специализированные платформы. На данный момент есть два способа получать эти данные: первый – это выгрузка с сайта ЕИС Закупки, второй – парсить эти данные с сайта, так как удобного API еще нет.

3. Обработка и предобработка данных. Собранные данные требуют тщательной обработки. На этом этапе осуществляется очистка данных от ошибок, дубликатов и некорректной информации. Также может потребоваться нормализация данных, чтобы привести их к единому формату, что облегчит дальнейший анализ. Сейчас в изначальном виде данные еще не подходят для работы и требуют значительной обработки.

4. Выбор и обучение моделей. На этом этапе необходимо выбрать подходящие алгоритмы машинного обучения для анализа данных. В зависимости от поставленных задач могут использоваться методы классификации (например, для определения надежности контрагентов) и регрессии (для прогнозирования вероятности успешного выполнения контракта). На данном этапе применяются различные методы в зависимости от подхода, выбранного в п.1 данного списка.

5. Валидация и тестирование. После обучения моделей необходимо провести валидацию и тестирование. Это позволяет оценить их точность и эффективность. Использование тестовых наборов данных, которые не использовались в процессе обучения, поможет выявить слабые места и настроить модели для повышения их предсказательной способности. Необходимо будет тестировать результаты на всей системе прогнозирования, а не сравнивать эти методы отдельно для данного модуля, так как итоговой целью является именно система прогнозирования исполнения государственных контрактов, а не его отдельного модуля.

6. Разработка пользовательского интерфейса. Важно создать удобный и интуитивно понятный интерфейс, который позволит пользователям легко

взаимодействовать с системой. Дашборды и визуализация данных должны обеспечивать возможность быстрого анализа и интерпретации результатов. Включение инструментов для генерации отчетов также повысит функциональность модуля.

7. Внедрение и обучение пользователей. После завершения разработки и тестирования системы следует этап внедрения. Важно обеспечить обучение пользователей, чтобы они могли эффективно использовать новую систему в своей работе. Поддержка и обратная связь от пользователей помогут в дальнейшем улучшении и адаптации системы к изменяющимся требованиям.

8. Мониторинг и обновление. После внедрения системы необходимо организовать постоянный мониторинг ее работы. Это позволит своевременно выявлять проблемы и недочеты, а также обновлять модели и алгоритмы по мере поступления новых данных. Регулярные обновления обеспечат актуальность системы и ее способность адаптироваться к изменениям в законодательстве и практике закупок.

Математические модели

В рамках данной работы были использованы такие методы как: логистическая регрессия, случайный лес, деревья решений, нейронные сети и методы кластеризации.

Логистическая регрессия. Была использована для определения вероятности того, что контрагент выполнит свои обязательства или нет. Эта модель проста в интерпретации и позволяет оценивать влияние различных факторов на результат. $X = \{x_1, x_2, \dots, x_n\}$ - множество параметров, характеризующих контрагента, количество выполненных и невыполненных контрактов, сумма выполненных контрактов, максимальный и минимальная цена полученного контракта и другие. Тогда

$$P = \frac{1}{1 + e^{-y}}, \quad (1)$$

где P – вероятность того, что контрагент выполнит контракт; $y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$, с весами $\{b_1, b_2, \dots, b_n\}$.

Случайный лес – ансамблевая модель, состоящая из множества деревьев решений, которая снижает риск переобучения и улучшает точность прогнозов. Эффективно работает с большими наборами данных и может обрабатывать как числовые, так и категориальные признаки, что очень эффективно при анализе таких объектов как контрагенты исполнения контракта. После многократного обучения n одинаковых базовых моделей на бутстрапированных выборках, полученный ансамбль производит конечный прогноз, просто агрегируя прогнозы базовых моделей. В случае классификации функция агрегирования представляет собой статистическую моду,

$$\hat{C}_{rf}^B(x) = \text{majority vote}[\hat{C}_b(x)]_1^B, \quad (2)$$

где $\hat{C}_b(x)$ – спрогнозированный класс b -го бутстрапированного дерева в случае классификации.

Методом кластеризации выбран K -средних. Он был использован для сегментации контрагентов на основе их характеристик, что помогло выделить группы с похожими рисками и поведением. В данном случае принадлежность контрагента к тому или иному сегменту был использован как параметр в итоговой системе прогнозирования исполнения государственных контрактов.

Нейронные сети показали низкие результаты по двум причинам. Первая – нейросеть не «видит» полную картину и не взаимодействует с остальными данными из данной системы, предыдущие методы показали более высокий результат по сравнению с ней. Вторая причина – достаточно небольшое количество данных ввиду конфиденциальности многих из них и их сложного получения.

Программный комплекс для анализа контрагентов

В рамках исследования был разработан практико-ориентированный комплекс программ, позволяющий проводить анализ контрагентов. Программный комплекс включает в себя ряд инструментов и модулей, которые совместно обеспечивают полноценный анализ контрагентов, их репутации и финансовых показателей. В состав программного комплекса входят следующие функции: автоматизированный сбор и интеграция информации из открытых источников (реестры, специализированные базы данных, новости и т. д.), обработка данных о контрагентах, их истории выполнения контрактов, финансовых показателей.

В ходе исследования программно было реализованы математические модели, описанные в предыдущем разделе. Результаты проведения серии вычислительных экспериментов с использованием математических моделей и комплекса программ приведены в следующем разделе.

Фрагмент программного комплекса, реализованного на языке Python представлен на рис. 2.

В программном комплексе используются алгоритмы машинного обучения для прогнозирования вероятности выполнения контракта по конкретным контрагентам на основе исторических данных, с визуализацией прогнозов в удобном формате (графики, таблицы), что позволяет анализировать данные и принимать обоснованные решения.

Интерфейс (рис. 3) позволяет пользователям управлять данными и получать необходимые отчеты, при этом обеспечивается высокий уровень защиты данных, что особенно важно при работе с конфиденциальной информацией. Использование данного комплекса помогает снизить риски, оптимизировать процессы принятия решений и сэкономить время благодаря автоматизации процессов сбора и анализа данных, что в итоге позволяет государственным учреждениям и организациям, участвующим в

государственных закупках, более эффективно управлять своими контрактами и взаимодействовать с контрагентами, повышая общую надежность и качество исполнения государственных обязательств.

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
import matplotlib.pyplot as plt
import seaborn as sns

# 1. Сбор данных
def load_data(file_path):
    df = pd.read_excel(file_path)
    return df

# 2. Анализ данных
def analyze_data(df):
    # Пример статистического анализа
    print(df.describe())
    return df.corr()

# 3. Подготовка данных для машинного обучения
def prepare_data(df):
    X = df.drop(columns=["target"]) # "target" - целевая переменная
    y = df["target"]
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
    return X_train, X_test, y_train, y_test

# 4. Прогнозирование
def train_model(X_train, y_train):
    model = RandomForestClassifier()
    model.fit(X_train, y_train)
    return model

def predict(model, X_test):
    return model.predict(X_test)

# 5. Визуализация результатов
def visualize_results(correlation_matrix):
    plt.figure(figsize=(10, 8))
    sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap='coolwarm')
    plt.title('Correlation Matrix')
    plt.show()
```

Рисунок 2 – Фрагмент программного комплекса, реализованного на языке Python

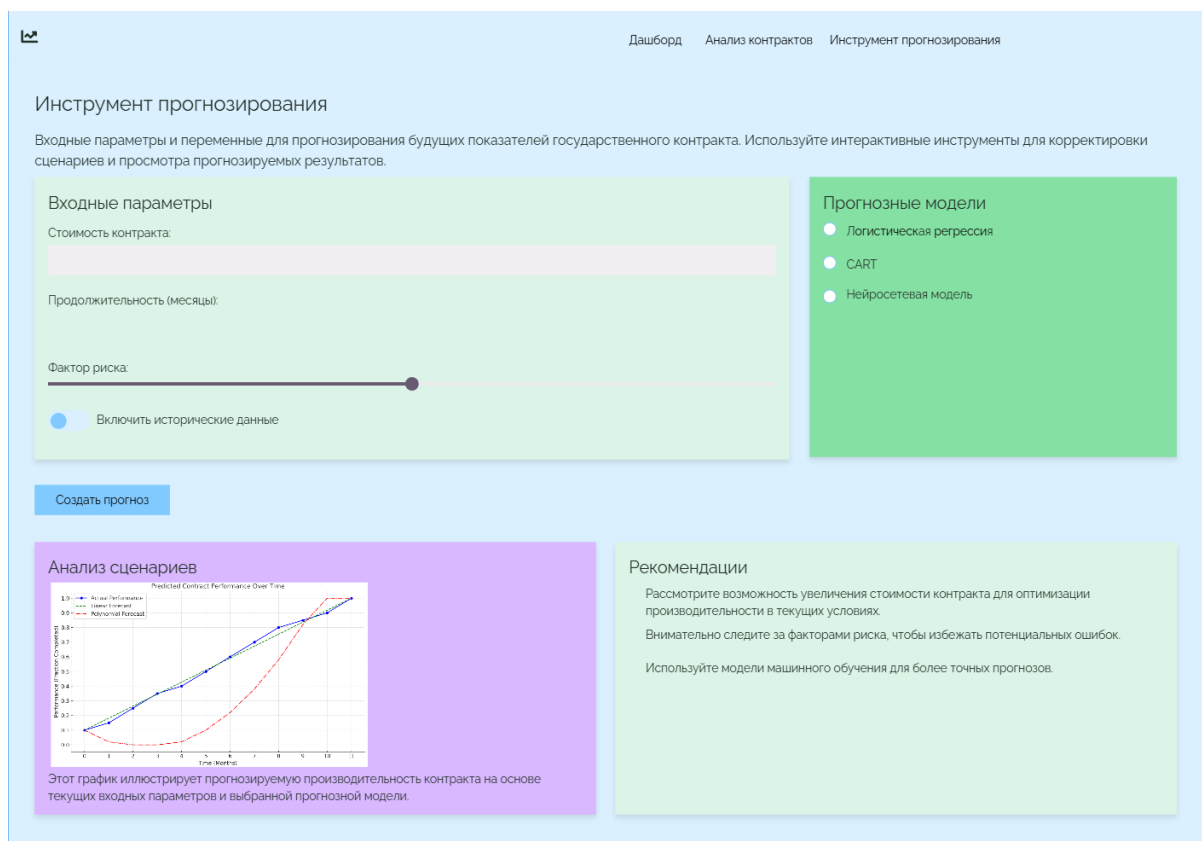


Рисунок 3 – Интерфейс программного комплекса

Результаты вычислительного эксперимента

В ходе подсчета метрики использовалась точность оценки вероятности того, что контрагент выполнит контракт. Для составления показателя использовался следующий подсчёт:

1. Контракт может завершиться тремя исходами:
 - а. Исполнение завершено;
 - б. Исполнение прекращено;
 - с. Аннулировано.
2. Каждому заказчику система подбирает результат, с которым может завершиться контракт (для подсчета логистической регрессии исход б и с считался единым).
3. Далее считалась точность оценки моделей.

В таблице № 1 приводится оценка точности и риски реализованных моделей.

Таблица № 1.

Метрики реализуемых моделей.

Используемая модель	Точность оценки	Риски
Дерево решений	97,89%	Тяжело определить вероятность исполнения в связи с бинарным ответом.
К-средних	97,88%	Распределение по всему n-мерной плоскости.
Логистическая регрессия	97,01%	Показатель исполнения либо сильно завышен, либо занижен.
Нейросетевая модель	87,65%	Малый набор данных

Заключение

Для максимальной эффективности модуль анализа контрагентов должен интегрироваться с другими системами в рамках управления закупками. Это позволит обмениваться данными и создавать более полное представление о каждом контрагенте. Например, интеграция с системами управления проектами и финансового учета даст возможность учитывать не только внешние факторы, но и внутренние характеристики организации. Такой подход создает единую экосистему, где информация о контрагентах может быть использована для принятия более обоснованных решений на всех уровнях. Это может включать автоматическую генерацию рекомендаций по выбору контрагентов, а также уведомления о возможных рисках и необходимости дополнительного анализа.

С развитием технологий и увеличением объемов данных, доступных для анализа, модуль анализа контрагентов будет продолжать

эволюционировать. Внедрение новых методов и инструментов, таких как обработка естественного языка и нейросетевые модели, позволит более эффективно анализировать отзывы, комментарии и другую текстовую информацию, связанную с контрагентами.

Кроме того, использование блокчейна может повысить уровень доверия к данным, обеспечивая их неизменность и доступность. Это позволит создавать более прозрачные и безопасные условия для проведения закупок, минимизируя риск мошенничества и злоупотреблений, что может послужить перспективным направлением развития данной проблематики. Данная разработка носит значимый прикладной характер и может быть полезна в различных отраслях экономики, где происходят государственные закупки. Расширение базы данных и совершенствование моделей прогнозирования исполнения контрактов способны повысить точность классификации контрагентов и принятия решений. Таким образом, дальнейшее развитие и применение предложенного программного комплекса может принести значительную пользу в сфере государственного и муниципального управления, способствуя оптимизации расходования бюджетных средств и повышению качества закупочной деятельности.

Практическая значимость данной работы состоит в возможности анализировать контракты разного рода. Результаты могут быть использованы для прогнозирования исполнения контрактов, в том числе в коммерческой сфере (отдельными предприятиями). Описанный метод и совокупность моделей будет полезна разработчикам в области программного обеспечения для мониторинга и прогнозирования закупочной деятельности.

Литература

1. Лукин В. К., Логинова М. М., Логинова В. Е. Формирование взаимосвязанного комплекса организационных инструментов обслуживания

финансовых программ // Экономика. Право. Печать. Вестник КСЭИ. 2013. №. 4. С. 245-249.

2. Федорович В. А., Муравник В. Б., Бочкарев О. И. США: военная экономика (организация и управление) // М.: Международные отношения. – 2013. – С. 36-46.

3. Цветков В. Корпоративный бизнес. Теория и практика. Litres, 2022. – С. 477-480.

4. McAfee R. P., McMillan J. Bidding for contracts: a principal-agent analysis // The RAND Journal of Economics. 1986. pp. 326-338.

5. Storm J. R. Outsourcing intelligence analysis: legal and policy risks // J. Nat'l Sec. L. & Pol'y. 2017. V. 9. P. 669.

6. Шевченко К. Ш. Особенности государственного контракта в сфере государственных закупок // Наука и образование: история и современность: электронный сборник материалов 74–75 внутривузовских научно-практических конференций. (Министерство науки и высшего образования Российской Федерации филиал Южно-Уральского государственного университета (НИУ) в г. Нижневартовске 19–20 апреля 2022 года). Издательство: Издательский центр ЮУрГУ". 2022. С. 89-93.

7. Проект «Национальный Рейтинг Прозрачности Закупок». URL: nprz.ru (дата обращения: 21.10.2024).

8. Корчагин С. А., Догадина Е. П., Мелентьев В. В., Никитин П. В., Сердечный Д. В. Автоматизированная система выдачи банковских гарантий на основе прогнозирования исполнения государственных контрактов // Инженерный вестник Дона. 2023. № 8 (104). URL: ivdon.ru/magazine/archive/n8y2023/8600/.

9. Шустова Е. С., Корчагин С. А., Феклин В. Г., Мелентьев В. В., Сердечный Д. В. Система поддержки принятия решений по выдаче

банковских гарантий с использованием методов машинного обучения // Современные наукоемкие технологии. 2023. № 7. С. 41–47.

10. Корчагин, С. А., Рубцов, Д. Ю., Сердечный, Д. В., Беспалова, Н. В. Применение моделей машинного обучения для прогнозирования исполнения государственных контрактов // Инженерный вестник Дона. 2023. № 9. URL: ivdon.ru/magazine/archive/n9y2024/9507/.

References

1. Lukin V. K., Loginova M. M., Loginova V. E. E`konomika. Pravo. Pechat`. Vestnik KSE`I. 2013. №. 4. pp. 245-249.

2. Fedorovich V. A., Muravnik V. B., Bochkarev O. I. SShA: voennaya e`konomika (organizaciya i upravlenie) [USA: military economy (organization and management)]. M.: Mezhdunarodny`e otnosheniya. 2013. pp. 36-46.

3. Czvetkov V. Korporativny`j biznes. Teoriya i praktika [Corporate business. Theory and practice.]. Litres, 2022. pp. 477-480.

4. McAfee R. P., McMillan J. The RAND Journal of Economics. 1986. pp. 326-338.

5. Storm J. R. Outsourcing intelligence analysis: legal and policy risks. J. Nat'l Sec. L. & Pol'y. 2017. V. 9. pp. 669.

6. Shevchenko K. Sh. Osobennosti gosudarstvennogo kontrakta v sfere gosudarstvenny`x zakupok [Features of the government contract in the field of public procurement]. Nauka i obrazovanie: istoriya i sovremennost`: e`lektronny`j sbornik materialov 74–75 vnutrivuzovskix nauchno-prakticheskix konferencij. (Ministerstvo nauki i vy`sshego obrazovaniya Rossijskoj Federacii filial Yuzhno-Ural`skogo gosudarstvennogo universiteta (NIU) v g. Nizhnevartovske 19–20 aprelya 2022 goda). Izdatel`stvo: Izdatel`skij centr YuUrGU". 2022. pp. 89-93.

7. Proekt «Nacional`ny`j Rejting Prozhachnosti Zakupok» [Project “National Procurement Transparency Rating”]. URL: nrpz.ru (date assessed: 21.10.2024).



8. Korchagin S. A., Dogadina E. P., Melent`ev V. V., Nikitin P. V., Serdechnyj D. V. Inzhenernyj vestnik Dona. 2023. № 8. URL: ivdon.ru/magazine/archive/n8y2023/8600/.

9. Shustova E. S., Korchagin S. A., Feklin V. G., Melent`ev V. V., Serdechnyj D. V. Sovremennyy`e naukoemkie texnologii. 2023. № 7. pp. 41–47.

10. Korchagin, S. A., Rubczov, D. Yu., Serdechny`j, D. V., Bespalova, N. V. Inzhenernyj vestnik Dona. 2023. № 9. URL: ivdon.ru/magazine/archive/n9y2024/9507/.

Дата поступления: 19.11.2024

Дата публикации: 2.01.2025