

# Построение регрессионной модели и модели классификации исполнения государственных контрактов

Н. А. Андриянов, Л. А. Ежова

<sup>1</sup>ФГБОУ ВО Финансовый университет при Правительстве Российской Федерации

*В данной работе рассматривается применение регрессионных и классификационных моделей для анализа и прогнозирования условий и рисков государственных контрактов. Авторы исследуют различные методы, включая логистическую регрессию, метод опорных векторов, а также градиентные бустинги, и оценивают их эффективность на реальных данных о контрактах в различных областях. Было показано, что модели машинного обучения могут быть эффективным инструментом для прогнозирования характеристик контрактов и оценки рисков.*

## Введение

В современном мире государственные контракты играют ключевую роль в регулировании экономических отношений и обеспечении выполнения государственных функций. Предсказание условий и рисков таких контрактов является важной задачей как для государственных органов, так и для коммерческих структур. В последние годы растет интерес к применению методов машинного обучения, особенно регрессионных и классификационных моделей, для анализа и прогнозирования параметров государственных контрактов.

## Построение регрессионной модели и модели классификации исполнения государственных контрактов

Регрессионные модели являются одним из наиболее распространенных методов прогнозирования условий государственных контрактов. Они позволяют определить связь между различными переменными и предсказать будущие значения одной переменной на основе значений других переменных.

В контексте государственных контрактов регрессионные модели могут использоваться для прогнозирования стоимости контрактов, длительности их выполнения, а также других параметров, которые могут быть важны для различных сторон.

Для создания регрессионной модели необходимо определить зависимости между независимыми и зависимой переменными. Независимые переменные могут включать такие параметры, как объем работ, наличие подрячков и т.д. Зависимая переменная может быть любым показателем контракта, который нужно предсказать.

После определения зависимостей и подбора модели, можно использовать ее для прогнозирования будущих значений зависимой переменной на основе значений независимых переменных. Регрессионные модели могут также помочь выявить факторы, которые могут влиять на изменение цены или других параметров контракта.

Однако стоит помнить, что регрессионные модели не могут предсказать будущее с точностью до единицы и могут содержать ошибки, которые необходимо учитывать при принятии решений на основе их прогнозов. Также важно учитывать, что регрессионные модели требуют точных данных, поэтому необходимо обеспечить правильную сборку, обработку и анализ данных перед созданием модели.

Модели классификации также могут быть использованы для анализа государственных контрактов. Они могут помочь в определении, отвечает ли определенный контракт заданным критериям, или отнести контракт к определенной категории или группе.

Один из примеров использования моделей классификации для государственных контрактов — это выявление несоответствий в заявках на участие в тендерах. Модели могут анализировать текстовые описания услуг или товаров, предлагаемых в заявке, и сравнивать их с критериями, установленными для данного контракта. Например, если контракт требует поставки определенных видов оборудования, модель может быстро определить, соответствует ли заявленное оборудование этим требованиям.

Кроме того, модели классификации могут использоваться для определения рисков в связи с контрактами. Они могут выявлять контракты, которые могут быть связаны с большими рисками и помогать в принятии решения о том, стоит ли принимать участие в тендере или заключать контракт.

Важно отметить, что модели классификации также требуют точных данных для обучения и принятия решений, и результаты моделей могут зависеть от того, какие данные были использованы для обучения. Поэтому важно заботиться о качестве данных и подходить к их сбору и обработке с особой внимательностью.

Работа посвящена прогнозированию «хороших» и «плохих» контрактов. Фактически, автором решается задача бинарной классификации. При этом для обучения и тестирования подготовлена выборка из порядка 12 тысяч записей. Автором было предложено разделить признаки на характерные для поставщика, заказчика и для самого контракта. Это позволило организовать более четкую структуру. Стоит отметить, что анализировались данные только среди государственных контрактов Ярославской области. Для решения задачи классификации в были обучены модели логистической регрессии, машины опорных векторов, модель классического градиентного бустинга и модель категориального бу-

стинга CatBoost . Наилучшие результаты (94,6%) показала модель категориального бустинга. Однако следует отметить, что в работе рассматривался довольно узкий подход, а при обработке реальных данных соотношение рискованных и нерискованных контрактов не обязательно будет 50% на 50%.

В другом исследовании рассматриваются модели прогнозирования рисков по контрактам в строительной области . То есть специфика задачи изначально была сужена автором. Автор использует XML-файлы из Единой информационной системы, а также анализируют заранее подготовленный датасет из недобросовестных поставщиков (сформированный в виде реестра недобросовестных организаций). Для определения успешности контрактов используются модели дефолта заемщика . В дальнейшем строится ряд моделей машинного обучения: логистическая регрессия, дерево решений и случайный лес. По критерию `roc_auc_score` (площадь под кривой) наилучшие результаты показала модель случайного леса. При этом значение метрики составило 0,8884.

В работе не до конца можно согласиться с экспертными оценками целевой переменной, однако было показано, что применение моделей машинного обучения для целей бинарной классификации может быть весьма целесообразно.

Исследование также посвящено оценке рисков при исполнении государственных контрактов . Новизной данной работы является дополнительное применение методов лингвистического анализа текстовых данных о государственных закупках. Это позволяет извлечь дополнительную информацию для модели прогнозирования. Для кодирования текстов используется мешок слов с TF-IDF. Также была попытка авторами спрогнозировать мультитэговые категории контрактов с использованием сверточных и рекуррентных искусственных нейронных сетей, которые привели к получению метрики F1-score около 60%.

Зарубежные исследователи также применяют технологии интеллектуального анализа данных для прогнозирования различных показателей по контрактам. Например, в работе с помощью регрессионной модели выполняется прогноз итоговой стоимости контракта . Автор использует логарифмическую стандартизацию данных, что обеспечивает им значение средней абсолютной относительной ошибки (MAPE) менее чем 8,5%.

Таким образом, в настоящее время исследователи и специалисты из разных областей проявляют большой интерес к моделям и методам регрессии и классификации при анализе государственных контрактов. В следующем разделе рассмотрим данные, для которых эти задачи будем решать далее.

#### **Описание данных о государственных контрактах**

Для разработки моделей регрессии и классификации государственных контрактов была подготовлена выгрузка данных с ЕИС. Информация была собрана в виде единого csv-файла, содержащего следующие поля:

- Номер реестровой записи контракта
- Заказчик: наименование
- Заказчик: ИНН
- Заказчик: КПП
- Уровень бюджета
- Источник финансирования контракта: наименование бюджета
- Источник финансирования контракта: наименование/вид внебюджетных средств
- Способ размещения заказа
- Номер извещения о проведении торгов
- Дата подведения результатов определения поставщика (подрядчика, исполнителя)
- Реквизиты документа, подтверждающего основание заключения контракта
- Контракт: дата
- Контракт: номер
- Предмет контракта
- Цена контракта
- Код бюджетной классификации
- КОСГУ
- КВР
- Идентификационный код закупки (ИКЗ)
- Объект закупки: наименование товаров, работ, услуг
- Объект закупки: код позиции
- Объект закупки: цена за единицу, рублей
- Объект закупки: количество поставленных товаров, выполненных работ, оказанных услуг
- Объект закупки: сумма, рублей
- Информация о поставщиках (исполнителях, подрядчиках) по контракту: наименование юридического лица (ф.и.о. физического лица)
- Информация о поставщиках (исполнителях, подрядчиках) по контракту: ИНН
- Информация о поставщиках (исполнителях, подрядчиках) по контракту: КПП
- Дата последнего изменения записи
- Дата исполнения контракта: по контракту

Пример представления данных приведен на рисунке 1.

Номер реестровой записи контракта	Заказчик: наименование	Заказчик: ИНН	Заказчик: КПП	Уровень бюджета	Источник финансирования контракта: наименование бюджета	Источник финансирования контракта: наименование/вид внебюджетных средств	Способ размещения заказа	Номер извещения о проведении торгов	Дата подведения результатов определения поставщика (подрядчика исполнителя)	Объект закупки: наименование товаров работ услуг	Объект закупки: код позиции	Объект закупки: цена за единицу рублей	Объект закупки: количество поставленных товаров выполненных работ оказанных услуг	Объект закупки: сумма рублей	Информация о поставщиках (исполнителях подрядчиках) по контракту: наименование юридического лица (ф.и.о. физического о лица)	Информация о поставщиках (исполнителях подрядчиках) по контракту: ИНН	Информация о поставщиках (исполнителях подрядчиках) по контракту: КПП	Дата последнего изменения записи	Дата исполнения контракта
0	МНЦИТАЛЬСКОЕ БУХГАЛТЕРСКОЕ УЧРЕЖДЕНИЕ КОПОРСКОГО РАЙОНА	504017308	50401001	НаИ	НаИ	Средства бюджетных учреждений	Закупка у единственного поставщика (подрядчика, исполнителя)	НаИ	...	Услуги по подготовке документации по закупке	38.00.20.100-00000001	33.20	70.004337349	2324.28	ГОСУДАРСТВЕННОЕ ИТАРНОЕ ПРЕДПРИЯТИЕ КОСГОСКО...	505408171	505401001	09.02.2023 14:52	01.12.2023
1	МНЦИТАЛЬСКОЕ БУХГАЛТЕРСКОЕ УЧРЕЖДЕНИЕ КОПОРСКОГО РАЙОНА	504017308	50401001	НаИ	НаИ	Средства бюджетных учреждений	Закупка у единственного поставщика (подрядчика, исполнителя)	НаИ	...	Услуги по подготовке документации по закупке	38.00.20.100-00000001	33.20	21.002710403	957.20	ГОСУДАРСТВЕННОЕ ИТАРНОЕ ПРЕДПРИЯТИЕ КОСГОСКО...	505408171	505401001	09.02.2023 14:52	01.12.2023
2	МНЦИТАЛЬСКОЕ БУХГАЛТЕРСКОЕ УЧРЕЖДЕНИЕ КОПОРСКОГО РАЙОНА	504017308	50401001	НаИ	НаИ	Средства бюджетных учреждений	Закупка у единственного поставщика (подрядчика, исполнителя)	НаИ	...	Услуги по водопользованию	37.00.10.000-00000002	31.43	21.001088542	1080.00	ГОСУДАРСТВЕННОЕ ИТАРНОЕ ПРЕДПРИЯТИЕ КОСГОСКО...	505408171	505401001	09.02.2023 14:52	01.12.2023
3	МНЦИТАЛЬСКОЕ БУХГАЛТЕРСКОЕ УЧРЕЖДЕНИЕ КОПОРСКОГО РАЙОНА	504017308	50401001	НаИ	НаИ	Средства бюджетных учреждений	Закупка у единственного поставщика (подрядчика, исполнителя)	НаИ	...	Услуги по водопользованию	37.00.10.000-00000002	31.43	70.002333269	9800.22	ГОСУДАРСТВЕННОЕ ИТАРНОЕ ПРЕДПРИЯТИЕ КОСГОСКО...	505408171	505401001	09.02.2023 14:52	01.12.2023
4	ФЕДЕРАЛЬНОЕ АГЕНСТВО НАУКИ И ТЕХНИЧЕСКОГО ПРОГРЕССА	001700203	001701001	Федеральный бюджет	Федеральный бюджет	НаИ	Закупка у единственного поставщика (подрядчика, исполнителя)	НаИ	...	водопользование	37.00.10.110	13.78	НаИ	НаИ	ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ "ВОД...	001705828	001701001	09.02.2023 14:52	01.01.2024

Рисунок 1 - Вывод информации о государственных контрактах  
Источник: составлено автором.

Ниже представлена информация о распределении всех признаков в собранном датасете и количестве полных данных.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20854 entries, 0 to 20853
Data columns (total 29 columns):
# Column                                Non-Null Count
Dtype
-----
0  Номер реестровой записи контракта 20854 non-null object
1  Заказчик: наименование 20854 non-null object
2  Заказчик: ИНН 20854 non-null object
3  Заказчик: КПП 20854 non-null object
4  Уровень бюджета 12442 non-null object
5  Источник финансирования контракта: наименование бюджета 12442 non-null object
6  Источник финансирования контракта: наименование/вид внебюджетных средств 8507 non-null object
7  Способ размещения заказа 20854 non-null object
8  Номер извещения о проведении торгов 15874 non-null object
9  Дата подведения результатов определения поставщика (подрядчика исполнителя) 16074 non-null object
10 Реквизиты документа подтверждающего основание заключения контракта 18057 non-null object
11 Контракт: дата 20854 non-null object
12 Контракт: номер 20835 non-null object
13 Предмет контракта 20725 non-null object
14 Цена контракта 20854 non-null object
15 Код бюджетной классификации 12155 non-null object
16 КОСГУ 12 non-null object
17 КВР 20829 non-null object
18 Идентификационный код закупки (ИКЗ) 20831 non-null object
19 Объект закупки: наименование товаров работ услуг 20854 non-null object
20 Объект закупки: код позиции 20854 non-null object
21 Объект закупки: цена за единицу рублей 20853 non-null object
22 Объект закупки: количество поставленных товаров выполненных работ оказанных услуг 9597 non-null object
23 Объект закупки: сумма рублей 9609 non-null object
24 Информация о поставщиках (исполнителях подрядчиках) по контракту: наименование юридического лица (ф.и.о. физического лица) 20854 non-null object
25 Информация о поставщиках (исполнителях подрядчиках) по контракту: ИНН 20854 non-null object
26 Информация о поставщиках (исполнителях подрядчиках) по контракту: КПП 20854 non-null object
27 Дата последнего изменения записи 20854 non-null object
28 Дата исполнения контракта: по контракту 20854 non-null object
dtypes: object(29)

```

memory usage: 4.6+ MB

Видно, что всего в датасете содержится 20854 контрактов. Также видим, что некоторые данные практически не заполнены. Например, значение столбца КОСГУ присутствует всего в 12 строках. Проведем удаление таких переменных в силу их бесполезности для моделей машинного обучения.

Далее посчитаем везде число дней до планового исполнения государственного контракта. Построим гистограмму для дней по исполнению контрактов. Она представлена на рисунке 2.

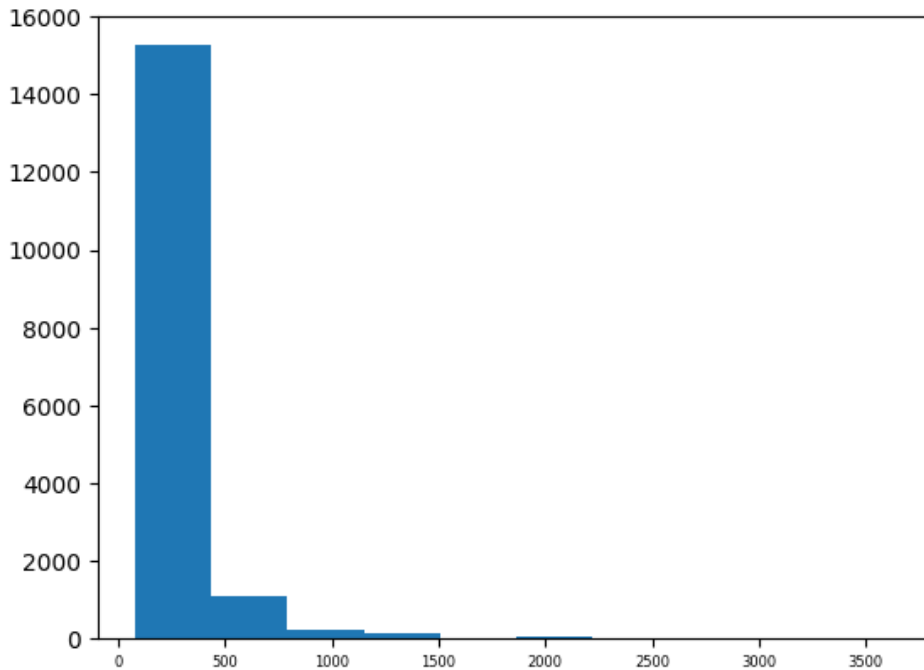


Рисунок 2 - Гистограмма сроков исполнения контрактов

Источник: составлено автором.

Анализ показывает, что в гистограмме присутствуют выбросные значения. Проверим это с помощью ящика с усами (рис. 3).

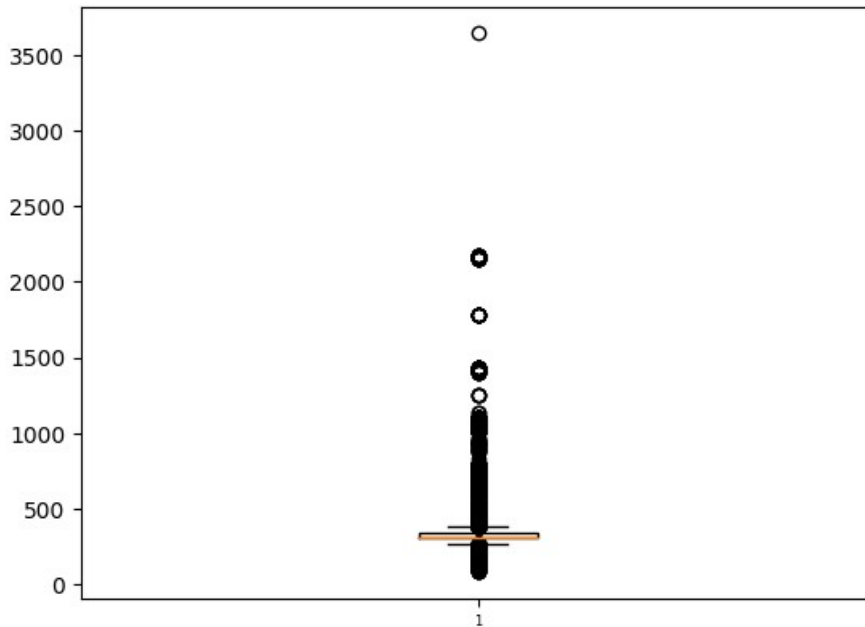


Рисунок 3 - Ящик с усами по срокам контрактов

Источник: составлено автором.

Отсечем данные по межквартильному размаху

$$IQR = 1,5 * (Q3 - Q1) \quad (1)$$

где Q1 – соответствует величине 25% процентиля (первый квартиль), а Q3 – величине 75% процентиля (третий квартиль).

Новая диаграмма принимает вид, представленный на рис. 4.

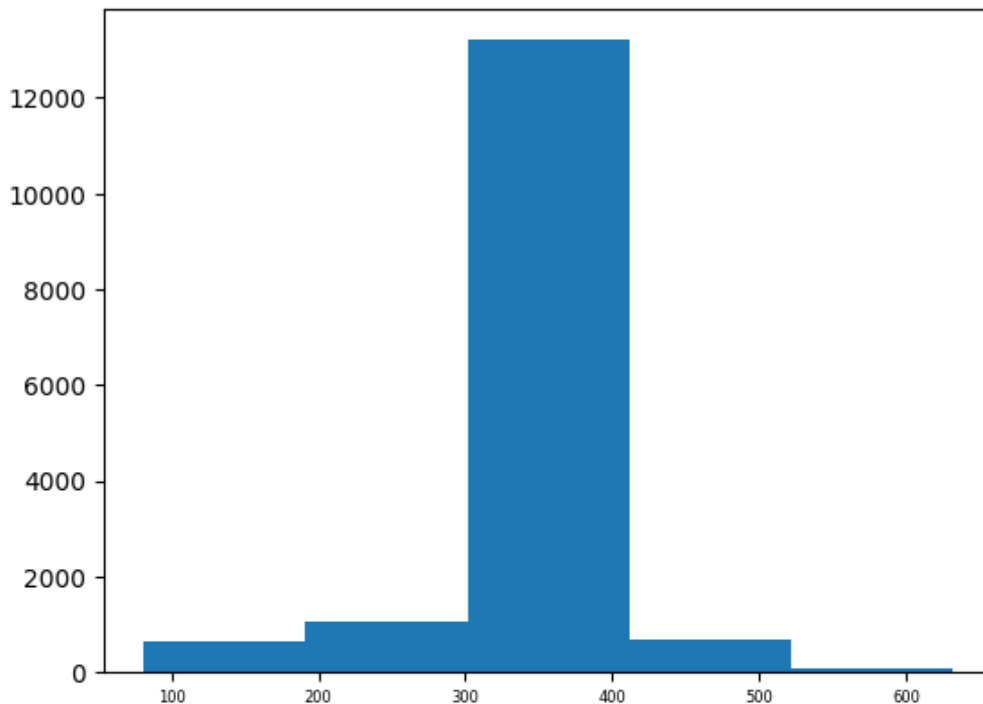


Рисунок 4 - Гистограмма сроков контрактов после предобработки

Источник: составлено автором.

Отметим, что мы отбросили все отрицательные дни до исполнения контрактов, т.е. там, где сроки были просрочены. В будущем, наоборот, можем использовать данную информацию для бинарной классификации контрактов.

Посмотрим распределения различных категориальных переменных. Данные представлены на рис. 5 – рис. 14.

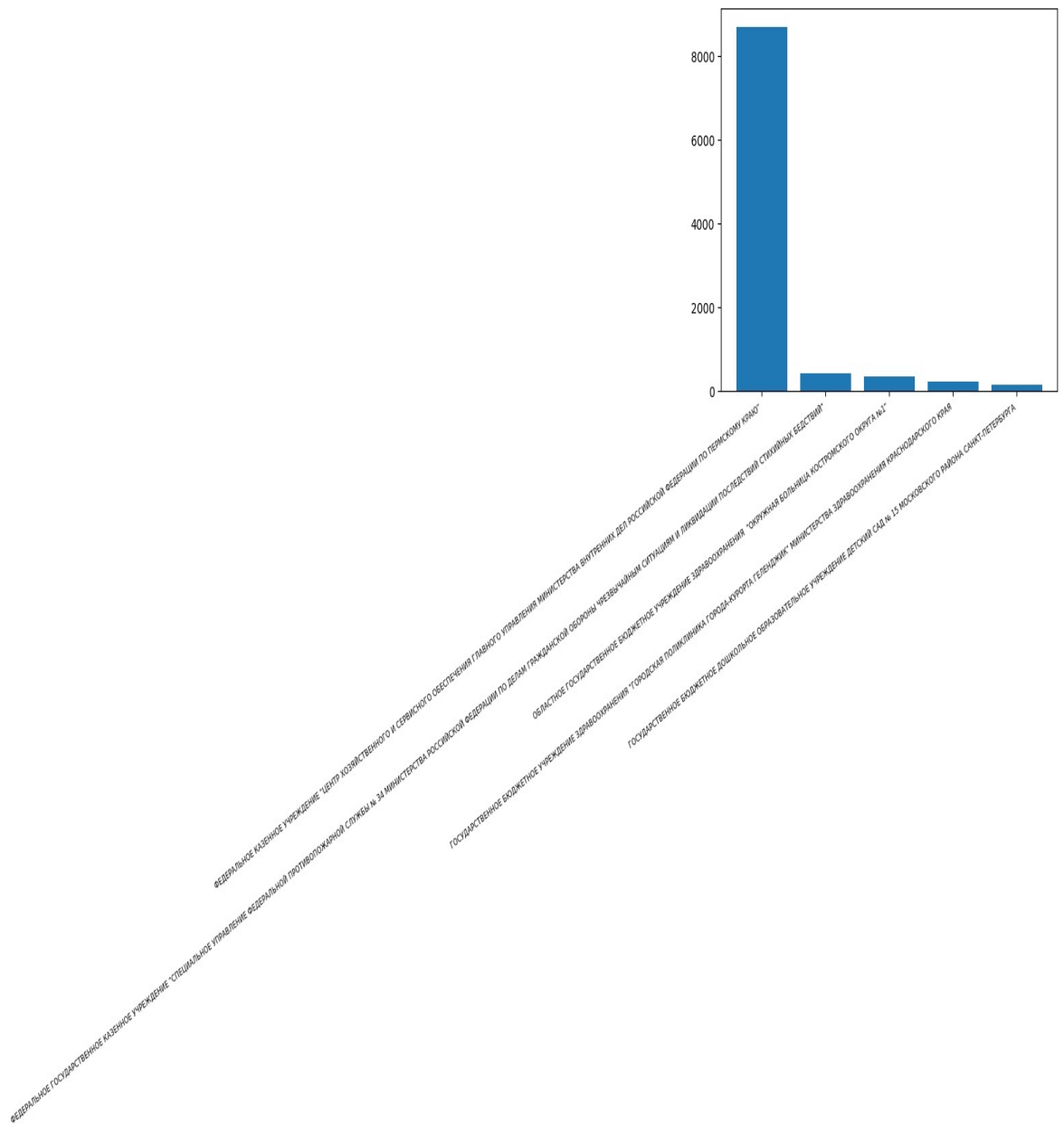


Рисунок 5 - Заказчик: наименование  
 Источник: составлено автором.

Видим, что в данных преобладает один заказчик.

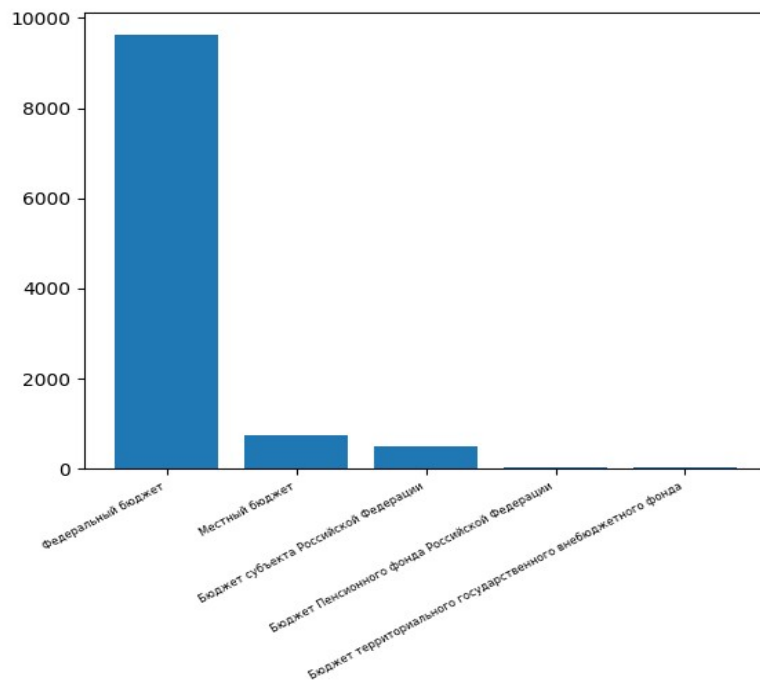


Рисунок 6 - Уровень бюджета  
 Источник: составлено автором.  
 Видим, что преобладает федеральный бюджет.

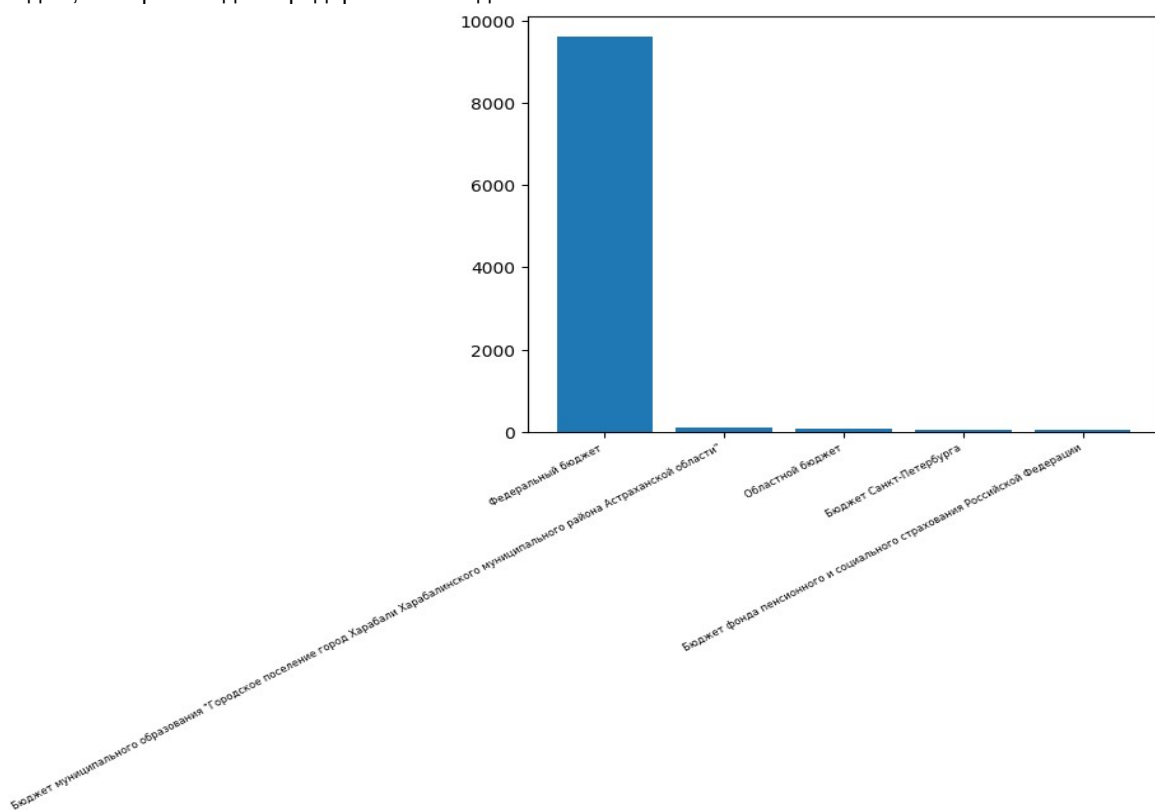


Рисунок 7 - Источник финансирования  
 Источник: составлено автором.  
 Также в финансировании преобладают средства федерального бюджета.

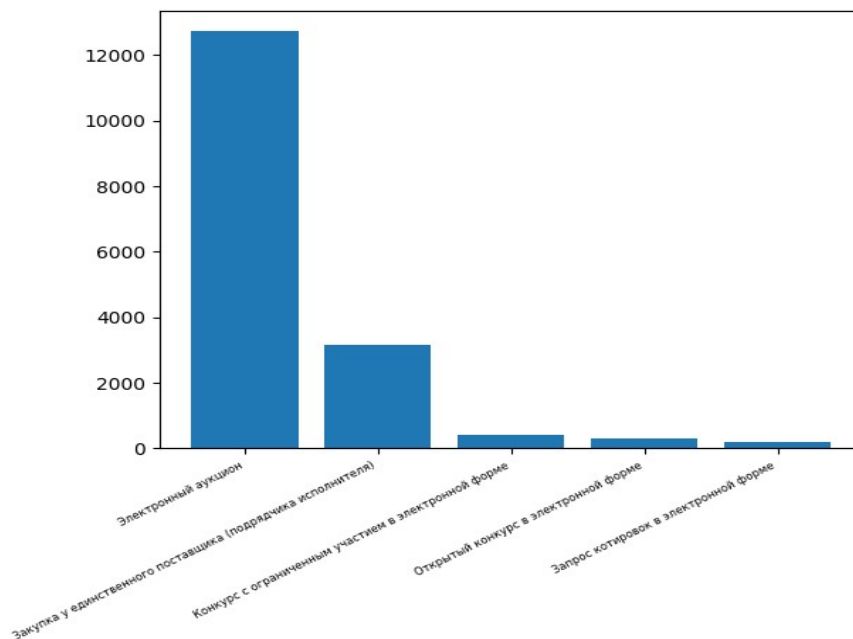


Рисунок 8 - Способ размещения заказа

Источник: составлено автором.

Основные способы размещения заказа – электронный аукцион и закупка у единственного поставщика.

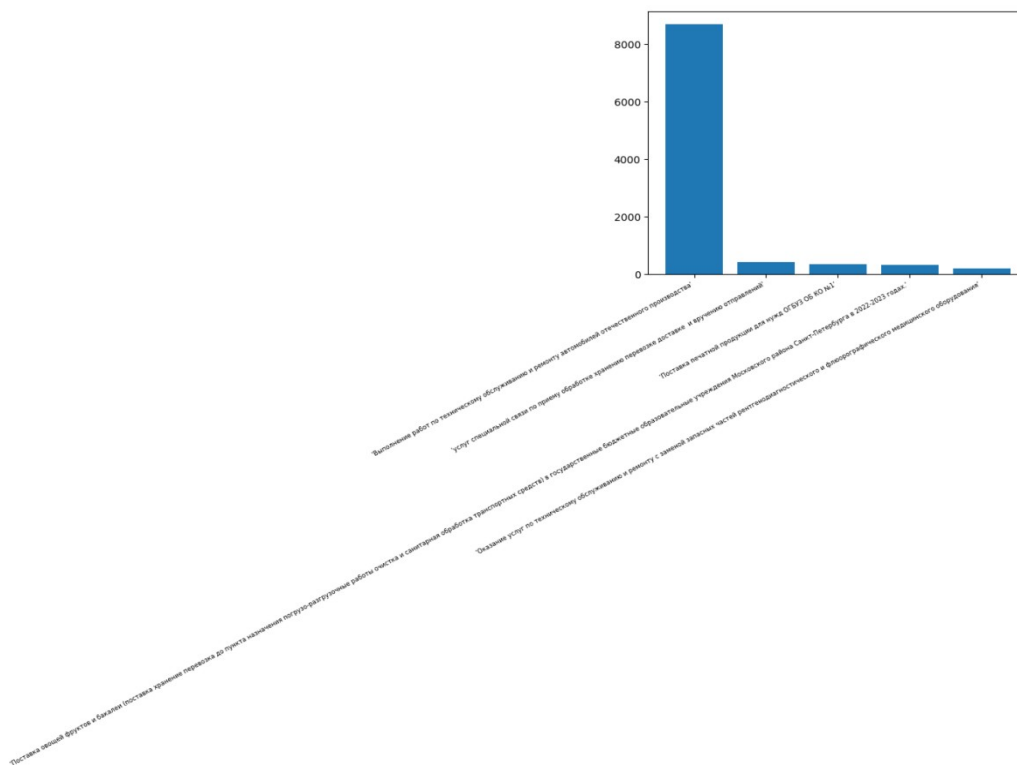


Рисунок 9 - Предмет контракта

Источник: составлено автором.

Предметом зачастую является техническое обслуживание и ремонт автомобилей (отечественных).



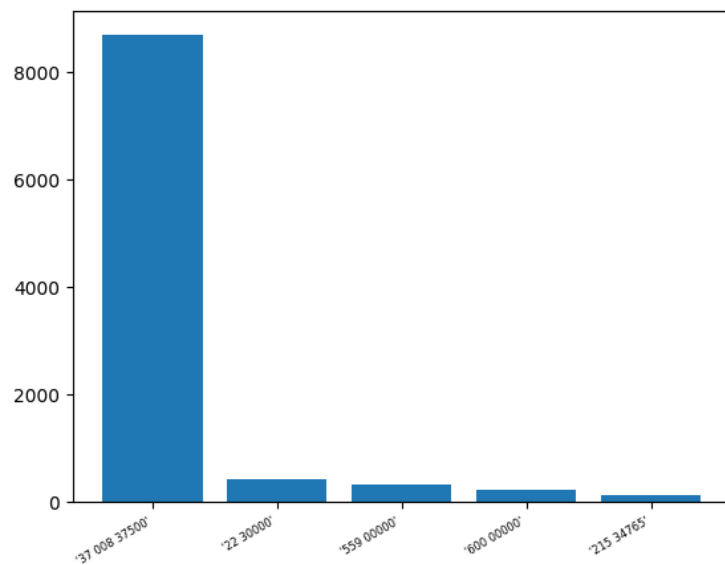


Рисунок 10 - Предмет контракта  
 Источник: составлено автором.

Интересно отметить, что цена контракта зачастую превосходит 3 миллиарда.

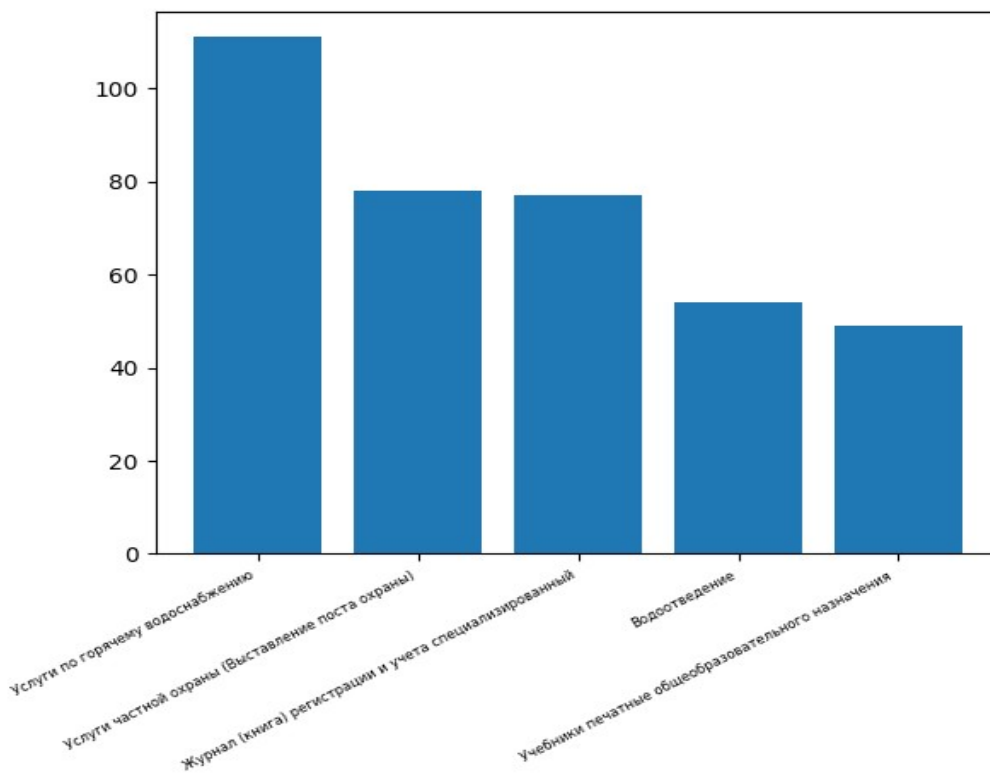


Рисунок 11 - Объект закупки  
 Источник: составлено автором.

Распределение объектов закупок более сбалансированное. Но преобладают услуги по горячему водоснабжению.

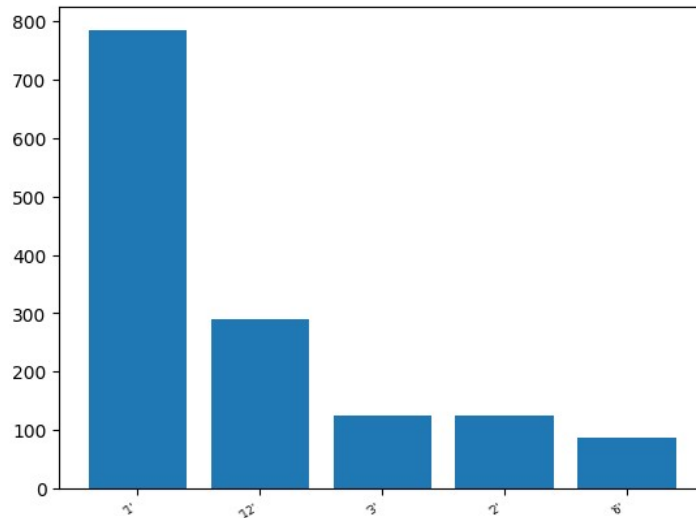


Рисунок 12 - Объект закупки: количество поставленных товаров выполненных работ оказанных услуг

Источник: составлено автором.

Как правило, поставляется один объект, однако часто встречаются и другие варианты.

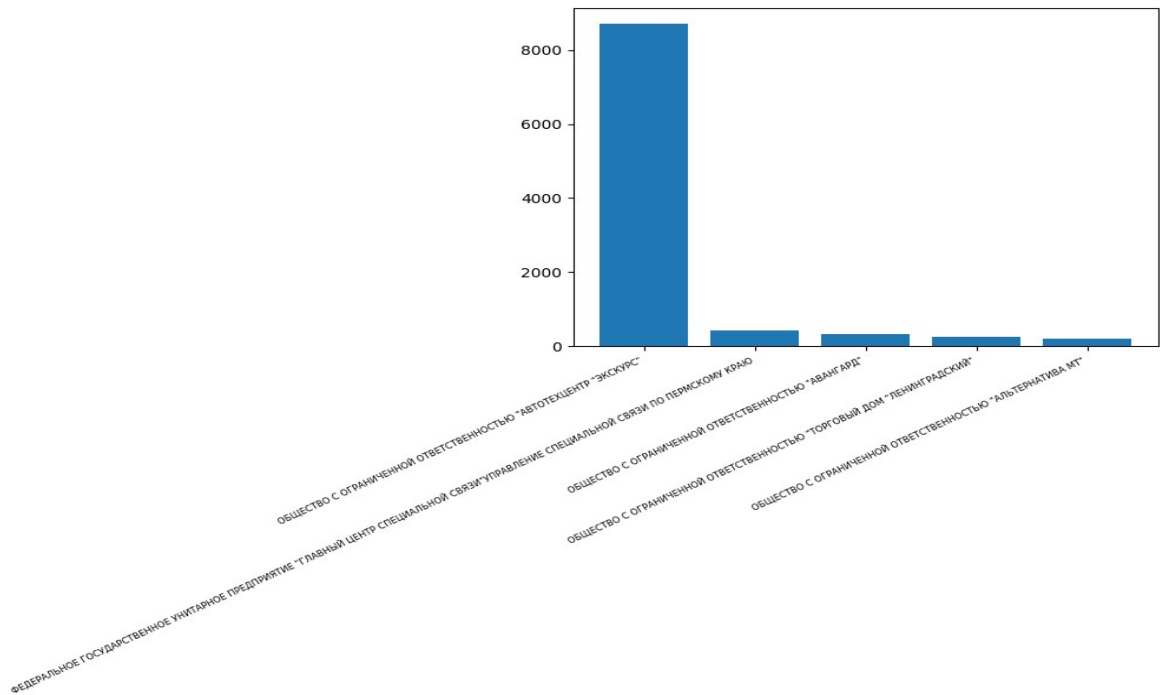


Рисунок 13 - Информация о поставщиках (исполнителях подрядчиках) по контракту: наименование юридического лица (ф.и.о. физического лица)

Источник: составлено автором.

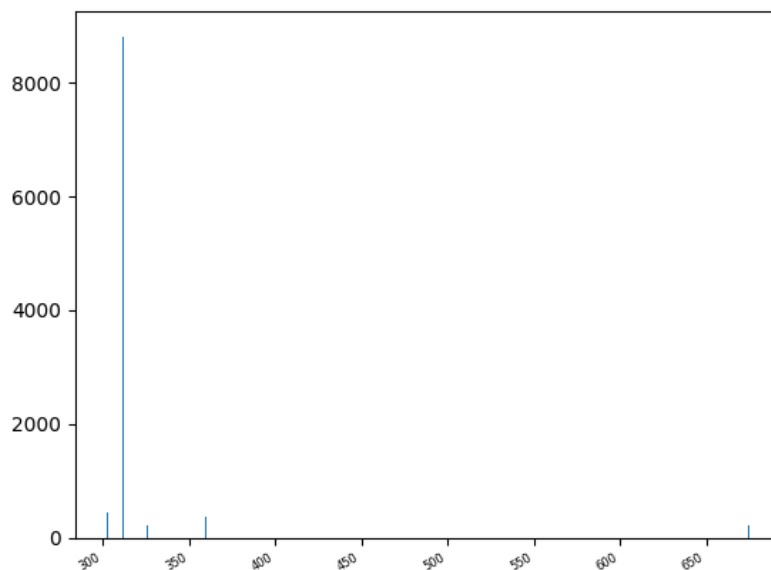


Рисунок 14 - Сроки выполнения контракта в днях  
Источник: составлено автором.

Очевидно, что сроки исполнения контрактов наиболее часто сосредоточены в промежутке около года.

Таким образом, можно сказать, что во многом собранный датасет преобладает значениями с сильным большинством, что может быть не информативно для модели, но и обладает также сбалансированными столбцами, которые могут нести больше информации.

#### **Заключение**

Исследование роли и применения регрессионных и классификационных моделей в анализе государственных контрактов подчеркивает их значимость в современном управлении государственными закупками. Результаты исследований показывают, что модели машинного обучения, такие как логистическая регрессия, метод опорных векторов и градиентные бустинги, могут быть успешно применены для прогнозирования стоимости, длительности выполнения и рисков государственных контрактов.

Однако следует учитывать, что эффективность моделей зависит от качества данных и правильного выбора признаков. Для достижения более точных прогнозов необходимо постоянно совершенствовать методы обработки и анализа данных, а также учитывать специфику отрасли и региональные особенности.

Таким образом, дальнейшие исследования в этой области могут включать разработку новых методов анализа данных, улучшение существующих моделей и расширение области применения машинного обучения для оптимизации государственных закупок и снижения рисков при выполнении государственных контрактов.

#### **Литература**

1. (Algorithms and methods for a risk management system in public procurement) (December 19, 2020). Available at SSRN: <https://ssrn.com/abstract=3792741> or <http://dx.doi.org/10.2139/ssrn.3792741>
2. Alsugair, A.M.; Alsanabani, N.M.; Al-Gahtani, K.S. Forecasting the Final Contract Cost on the Basis of the Owner's Cost Estimation Using an Artificial Neural Network. *Buildings* 2023, 13, 786. <https://doi.org/10.3390/buildings13030786>
3. Елисеев Д., Романов Д. Машинное обучение: прогнозирование рисков госзакупок // Открытые системы. СУБД, 2018 № 2 [Электронный журнал] Режим доступа: <https://www.osp.ru/os/2018/2/13054186> (дата обращения 08.05.2023)
4. Иванов Н.А., Валпетерс М.Л., Киреев И.А. «Большие данные» и машинное обучение при управлении рисками невыполнения обязательств по контрактам в строительной отрасли // Промышленное и гражданское строительство. 2019. № 5. С. 81 - 87.
5. Софронова В.В. Оценка дефолта заемщика // Финансовая аналитика: проблемы и решения. 2016. №3. С. 39-48.
6. Rey, Aleksey and Andronova, Ekaterina and Shatilov, A. A. and Gordeev, Dmitry and Shagarov, Dmitriy and Filippov, D. V., Алгоритмы и методы для системы управления рисками в государственных закупках

7. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* 521, 436–444 (2015). <https://doi.org/10.1038/nature14539>
8. Quinlan, J.R. Induction of decision trees. *Mach Learn* 1, 81–106 (1986). <https://doi.org/10.1007/BF00116251>
9. Бакрунов Юрий Октавьевич, and Васильева Елена Юрьевна. "ОЦЕНКА РИСКА ИННОВАЦИОННЫХ ПРОЕКТОВ С ПРИМЕНЕНИЕМ МЕТОДА ДЕРЕВА РЕШЕНИЙ" *Московский экономический журнал*, no. 5, 2022, pp. 353-365.
10. Cristianini, N., Ricci, E. (2008). Support Vector Machines. In: Kao, MY. (eds) *Encyclopedia of Algorithms*. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-30162-4\\_415](https://doi.org/10.1007/978-0-387-30162-4_415)
11. Михайлов И.С., Зеар Аунг, and Йе Тху Аунг. "РАЗРАБОТКА МОДИФИКАЦИИ МЕТОДА ОПОРНЫХ ВЕКТОРОВ ДЛЯ РЕШЕНИЯ ЗАДАЧИ КЛАССИФИКАЦИИ С ОГРАНИЧЕНИЯМИ НА ПРЕДМЕТНУЮ ОБЛАСТЬ" *Программные продукты и системы*, vol. 33, no. 3, 2020, pp. 439-448.
12. Cover, Thomas M.; Hart, Peter E. (1967). "Nearest neighbor pattern classification" (PDF). *IEEE Transactions on Information Theory*. 13 (1): 21–27. CiteSeerX 10.1.1.68.2616. doi:10.1109/TIT.1967.1053964. S2CID 5246200.
13. 7. Кернога А.Л., and Бурак Т.И.. "Сравнение подходов к прогнозированию методом ближайших соседей" *Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления*, no. 13, 2015, pp. 26-33.
14. Breiman, L. Random Forests. *Machine Learning* 45, 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
15. Jerome H. Friedman Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* Vol. 29, No. 5 (Oct., 2001), pp. 1189-1232 (44 pages)
16. Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin "Why Should I Trust You?": Explaining the Predictions of Any Classifier
17. Andriyanov, N. Methods for Preventing Visual Attacks in Convolutional Neural Networks Based on Data Discard and Dimensionality Reduction. *Appl. Sci.* 2021, 11, 5235. <https://doi.org/10.3390/app11115235>
18. Андриянов Н.А. АНАЛИЗ ЭФФЕКТИВНОСТИ МЕТОДОВ СОКРАЩЕНИЯ РАЗМЕРНОСТИ ДАННЫХ В ОБРАБОТКЕ ИЗОБРАЖЕНИЙ // *Современная математика и концепции инновационного математического образования*. 2021. Т. 8. № 1. С. 217-223.
19. Yu-Lin He, Xiao-Liang Zhang, Wei Ao, Joshua Zhexue Huang Determining the optimal temperature parameter for Softmax function in reinforcement learning // *Applied Soft Computing*, Volume 70, September 2018, Pages 80-85
20. Фаустова К.И.. "Нейронные сети: применение сегодня и перспективы развития" *Территория науки*, no. 4, 2017, pp. 83-87.

#### References in Cyrillics

1. (Algorithms and methods for a risk management system in public procurement) (December 19, 2020). Available at SSRN: <https://ssrn.com/abstract=3792741> or <http://dx.doi.org/10.2139/ssrn.3792741>
2. Alsugair, A.M.; Alsanabani, N.M.; Al-Gahtani, K.S. Forecasting the Final Contract Cost on the Basis of the Owner's Cost Estimation Using an Artificial Neural Network. *Buildings* 2023, 13, 786. <https://doi.org/10.3390/buildings13030786>
3. Eliseev D., Romanov D. Machine learning: forecasting the risks of government procurement // *Open systems. DBMS*, 2018 No. 2 [Electronic journal] Access mode: <https://www.osp.ru/os/2018/2/13054186> (access date 05/08/2023)
4. Ivanov N.A., Valpeters M.L., Kireev I.A. "Big data" and machine learning in managing the risks of non-fulfillment of obligations under contracts in the construction industry // *Industrial and civil construction*. 2019. No. 5. P. 81 - 87.
5. Sofronova V.V. Assessment of borrower default // *Financial analytics: problems and solutions*. 2016. No. 3. pp. 39-48.
6. Rey, Aleksey and Andronova, Ekaterina and Shatilov, A. A. and Gordeev, Dmitry and Shagarov, Dmitriy and Filippov, D. V., Algorithms and methods for a risk management system in public procurement
7. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* 521, 436–444 (2015). <https://doi.org/10.1038/nature14539>
8. Quinlan, J.R. Induction of decision trees. *Mach Learn* 1, 81–106 (1986). <https://doi.org/10.1007/BF00116251>
9. Bakrunov Yuri Oktavievich, and Vasilyeva Elena Yurievna. "RISK ASSESSMENT OF INNOVATION PROJECTS USING THE DECISION TREE METHOD" *Moscow Economic Journal*, no. 5, 2022, pp. 353-365.

10. Cristianini, N., Ricci, E. (2008). Support Vector Machines. In: Kao, M.Y. (eds) Encyclopedia of Algorithms. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-30162-4\\_415](https://doi.org/10.1007/978-0-387-30162-4_415)
11. Mikhailov I.S., Zear Aung, and Ye Thu Aung. "DEVELOPMENT OF A MODIFICATION OF THE SUPPORT VECTOR METHOD FOR SOLVING THE CLASSIFICATION PROBLEM WITH DOMAIN CONSTRAINTS" Software Products and Systems, vol. 33, no. 3, 2020, pp. 439-448.
12. Cover, Thomas M.; Hart, Peter E. (1967). "Nearest neighbor pattern classification" (PDF). IEEE Transactions on Information Theory. 13 (1): 21–27. CiteSeerX 10.1.1.68.2616. doi:10.1109/TIT.1967.1053964. S2CID 5246200.
13. Kernoga A.L., and Burak T.I.. "Comparison of approaches to forecasting using the nearest neighbor method" Bulletin of the Perm National Research Polytechnic University. Electrical engineering, information technology, control systems, no. 13, 2015, pp. 26-33.
14. Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
15. Jerome H. Friedman Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics Vol. 29, No. 5 (Oct., 2001), pp. 1189-1232 (44 pages)
16. Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin "Why Should I Trust You?": Explaining the Predictions of Any Classifier
17. Andriyanov, N. Methods for Preventing Visual Attacks in Convolutional Neural Networks Based on Data Discard and Dimensionality Reduction. Appl. Sci. 2021, 11, 5235. <https://doi.org/10.3390/app11115235>
18. Andriyanov N.A. ANALYSIS OF THE EFFECTIVENESS OF DATA DIMENSION REDUCTION METHODS IN IMAGE PROCESSING // Modern mathematics and concepts of innovative mathematical education. 2021. T. 8. No. 1. P. 217-223.
19. Yu-Lin He, Xiao-Liang Zhang, Wei Ao, Joshua Zhexue Huang Determining the optimal temperature parameter for Softmax function in reinforcement learning // Applied Soft Computing, Volume 70, September 2018, Pages 80-85
20. Faustova K.I.. "Neural networks: application today and development prospects" Territory of Science, no. 4, 2017, pp. 83-87.

#### **Ключевые слова**

Государственные контракты, машинное обучение, регрессионные модели, классификационные модели, прогнозирование, риски.

*Андрянов Никита Андреевич*

*Заведующий молодежной лабораторией компьютерного зрения Доцент Кафедры анализа данных и машинного обучения Финансового университета при Правительстве Российской Федерации  
[naandriyanov@fa.ru](mailto:naandriyanov@fa.ru)*

*Ежова Лилия Альбертовна*

*Аспирантка и ассистентка Кафедры бизнес информатики Факультета информационных технологий и анализа данных Финансового университета при Правительстве Российской Федерации  
[laezhova@fa.ru](mailto:laezhova@fa.ru)*

#### **Keywords**

Government contracts, machine learning, regression models, classification models, forecasting, risks.  
JELclassification 5.2.2 – Mathematical, statistical and instrumental methods in economics

#### **Abstract**

This paper discusses the use of regression and classification models for analyzing and predicting the conditions and risks of government contracts. The authors examine various methods, including logistic regression, support vector machines, and gradient boosting, and evaluate their performance on real contract data in various domains. It has been shown that machine learning models can be an effective tool for predicting contract performance and risk assessment.