

Применение моделей машинного обучения для прогнозирования исполнения государственных контрактов

С.А. Корчагин, Д.Ю. Рубцов, Д.В. Сердечный, Н.В. Беспалова

Финансовый университет при Правительстве Российской Федерации, Москва

Аннотация: В работе проанализированы существующие подходы к прогнозированию исполнения контрактов, включая традиционные статистические модели и современные методы на основе машинного обучения. Проведен сравнительный анализ различных алгоритмов машинного обучения, таких, как логистическая регрессия, деревья решений, случайный лес и нейронные сети, для выявления наиболее эффективных моделей прогнозирования. В качестве исходных данных использовалась обширная база информации о государственных контрактах, включающая информацию о подрядчиках, условиях контрактов, сроках исполнения и других значимых факторах. Разработан прототип интеллектуальной системы прогнозирования, проведено тестирование на реальных данных, а также оценка точности и надежности получаемых прогнозов. Результаты исследования показывают, что применение методов машинного обучения позволяет значительно повысить качество прогнозирования исполнения государственных контрактов по сравнению с традиционными подходами.

Ключевые слова: интеллектуальная система, математическое моделирование, государственные закупки, государственные контракты, программный комплекс, прогнозирование, машинное обучение.

Введение

Государственные закупки играют ключевую роль в реализации государственных программ и проектов, обеспечивая поставку необходимых товаров, работ и услуг [1,2]. Своевременное и качественное исполнение государственных контрактов является важным условием эффективности государственного управления [3,4]. Однако, на практике нередко возникают ситуации, когда государственные контракты не исполняются в полном объеме или с существенными нарушениями сроков. Это приводит к срыву реализации государственных инициатив, неэффективному расходованию бюджетных средств и снижению доверия к государственным институтам.

Для решения этой проблемы необходимы инструменты, позволяющие заблаговременно выявлять риски неисполнения или ненадлежащего исполнения государственных контрактов. Традиционные методы прогнозирования, основанные на статистическом анализе и экспертных

оценках, зачастую не обеспечивают требуемой точности и оперативности [5,6]. В этой связи все больший интерес вызывает применение методов искусственного интеллекта [7,8], в частности, моделей машинного обучения, для прогнозирования исполнения государственных контрактов.

Статья посвящена исследованию возможностей применения моделей машинного обучения для прогнозирования исполнения государственных контрактов. В работе проводится сравнительный анализ различных алгоритмов машинного обучения, оценивается их эффективность в задачах прогнозирования, а также описывается разработка прототипа интеллектуальной системы, способной заблаговременно выявлять риски срыва сроков и нарушения условий государственных контрактов. Результаты исследования демонстрируют, что использование методов искусственного интеллекта позволяет значительно повысить точность и оперативность прогнозирования исполнения государственных контрактов по сравнению с традиционными подходами. Предлагаемые решения могут быть эффективно внедрены в практику государственных закупок для повышения эффективности управления бюджетными средствами.

Модели машинного обучения

Для прогнозирования исполнения государственных контрактов в рамках данного исследования были рассмотрены различные модели машинного обучения. Выбор подходящих алгоритмов проводился с учетом особенностей задачи прогнозирования, характера имеющихся данных, а также требований к точности и интерпретируемости получаемых моделей.

Одним из базовых алгоритмов, применяемых в задачах бинарной классификации, является логистическая регрессия. Данная модель позволяет оценить вероятность наступления целевого события (в нашем случае - ненадлежащего исполнения государственного контракта) на основе линейной

комбинации входных признаков. Логистическая регрессия обладает высокой интерпретируемостью, что важно для объяснения полученных прогнозов. В контексте прогнозирования исполнения государственных контрактов логистическая регрессия может быть использована для оценки вероятности ненадлежащего исполнения контракта на основе имеющейся информации о его характеристиках.

Пусть имеется обучающая выборка из N объектов (государственных контрактов), каждый из которых описывается набором признаков $X = \{x_1, x_2, \dots, x_m\}$, где m - количество рассматриваемых признаков. Каждому объекту сопоставляется бинарная целевая переменная y , принимающая значение 1, если контракт был исполнен ненадлежащим образом, и 0 в противном случае. Логистическая регрессия моделирует вероятность принадлежности объекта к классу "ненадлежащее исполнение" как функцию линейной комбинации его признаков:

$$p(y=1|X) = 1 / (1 + \exp(-(w_0 + w_1x_1 + w_2x_2 + \dots + w_m * x_m))),$$

где w_0, w_1, \dots, w_m - коэффициенты (веса) модели, подлежащие оценке на основе обучающей выборки.

Для оценки весовых коэффициентов w применяется метод максимального правдоподобия. Целевая функция, максимизируемая в процессе обучения, имеет следующий вид:

$$L(w) = \sum y \cdot \log(p(y=1|X)) + (1-y) \cdot \log(1-p(y=1|X)),$$

где суммирование ведется по всем объектам обучающей выборки.

Оптимальные значения весов w находятся путем итеративной минимизации функции потерь $L(w)$ с использованием методов градиентного спуска или других оптимизационных алгоритмов.

Полученная модель логистической регрессии позволяет для любого нового объекта (государственного контракта) с известными признаками X вычислить вероятность ненадлежащего исполнения $p(y=1|X)$. Данная

вероятность может быть использована в качестве прогноза риска срыва исполнения контракта.

Важным преимуществом логистической регрессии является возможность интерпретации коэффициентов w , что позволяет выявить наиболее значимые факторы, влияющие на вероятность ненадлежащего исполнения контракта.

Другим популярным классом моделей машинного обучения являются деревья решений. Данный подход позволяет строить непараметрические модели, которые в виде иерархической структуры правил разбивают пространство входных признаков на области, соответствующие классам прогнозируемого события. Деревья решений отличаются высокой наглядностью и простотой интерпретации получаемых результатов. Формализованная модель деревьев решений для прогнозирования исполнения государственных контрактов может быть представлена следующим образом. Пусть $X = \{x_1, x_2, \dots, x_n\}$ - множество предикторов (характеристик контракта, заказчика, подрядчика, внешних факторов), y - целевая переменная (успешное исполнение контракта, да/нет). Алгоритм построения дерева решений можно описать в следующем виде:

1. Выбор признака для расщепления узла. Информационный критерий Шеннона: $I(X) = -\sum P(x) \log_2 P(x)$, где $P(x)$ - вероятность появления значения x в обучающей выборке. Информационный выигрыш для признака x_i :

$I \cdot G(x_i) = I(X) - \sum |X_j| / |X| \cdot I(X_j)$, где X_j - подмножество объектов, в которых признак x_i имеет значение j . Выбирается признак с максимальным информационным выигрышем: $x^* = \operatorname{argmax}(I \cdot G(x_i))$.

2. Расщепление узла по выбранному признаку: $X_j = \{x \in X \mid x_i = j\}$.

3. Рекурсивное построение поддеревьев для каждого подмножества X_j .

4. Остановка рекурсии, когда: все объекты в узле относятся к одному классу ($y = const$); достигнута максимальная глубина дерева; количество объектов в узле меньше заданного минимума.

5. Присвоение листовым узлам меток классов:

$y_{pred} = \operatorname{argmax}(P(y/X))$, где $P(y/X)$ - вероятность класса y при заданных признаках X .

Полученное дерево решений может быть использовано для прогнозирования успешности исполнения новых государственных контрактов. Для нового контракта с характеристиками X , вероятность успешного исполнения вычисляется как: $P(y=1/X) = P(y=1/x_1, x_2, \dots, x_n)$. Таким образом, модель деревьев решений позволяет выявить наиболее значимые факторы, влияющие на исполнение контрактов, и строить прогнозы на основе исторических данных.

Для повышения устойчивости и обобщающей способности моделей на основе деревьев решений применяется ансамблевый метод случайного леса. Он основан на обучении множества независимых деревьев решений на различных подвыборках обучающих данных и последующем объединении их прогнозов. Случайный лес показывает высокую точность предсказаний при сохранении интерпретируемости.

В рамках исследования была рассмотрена еще одна модель машинного обучения XGBoost (Extreme Gradient Boosting). Алгоритм XGBoost использует ансамблевый метод градиентного бустинга деревьев решений для построения модели. На каждой итерации t алгоритм строит новое дерево решений $h_t(x)$ и добавляет его к текущей модели $f_{t-1}(x)$: $f_t(x) = f_{t-1}(x) + \eta * h_t(x)$, где η - скорость обучения (learning rate). Целевая функция, которую XGBoost пытается минимизировать, имеет следующий вид: $L(y, f(x)) = \sum L(y(i), f(x(i))) + \Omega(f)$, где $\Omega(f)$ - функция регуляризации, которая

контролирует сложность модели и предотвращает переобучение. Функция регуляризации $\Omega(f)$ определяется как: $\Omega(f) = \gamma T + 1/2 \cdot \lambda \cdot \sum w_j^2$, где T - число листьев (терминальных узлов) в дереве, w_j - значение, хранимое в j -м листе, γ и λ - гиперпараметры регуляризации. На каждой итерации t алгоритм XGBoost решает следующую оптимизационную задачу: $(h_t, w_t) = \operatorname{argmin}_{h,w} \sum L(y^i, f_{t-1}(x(i)) + h(x^i)) + \Omega(h)$. Таким образом, XGBoost строит ансамбль деревьев решений, последовательно добавляя новые деревья, которые минимизируют целевую функцию с учетом регуляризации. Это позволяет получать высокоточные и устойчивые к переобучению модели.

Наряду с классическими алгоритмами машинного обучения, в данном исследовании также рассматривались модели на основе нейронных сетей. Нейронные сети способны автоматически извлекать сложные нелинейные закономерности в данных, что может быть полезно для прогнозирования исполнения контрактов. Однако нейронные сети зачастую характеризуются меньшей интерпретируемостью по сравнению с другими рассмотренными алгоритмами.

Выбор наиболее эффективной модели прогнозирования осуществлялся на основе сравнительного анализа точности, надежности и интерпретируемости рассмотренных алгоритмов на репрезентативных наборах данных о государственных контрактах. Результаты данного сравнения представлены в следующем разделе.

Результаты работы

В данном исследовании были применены различные модели машинного обучения для прогнозирования исполнения государственных контрактов. Были протестированы следующие модели: линейная регрессия, дерево решений, случайный лес и нейронная сеть. Для оценки качества моделей использовались следующие метрики: коэффициент детерминации

(R-squared), средняя абсолютная ошибка (MAE) и средняя квадратическая ошибка (RMSE). Модели были оценены с помощью кросс-валидации. Для обучения моделей использовались данные из открытых источников [9, 10].

Результаты показали, что модель XGBoost продемонстрировала наилучшие показатели качества с $R\text{-squared}=0.82$, $MAE=0.15$ и $RMSE=0.21$. Это говорит о том, что данная модель эффективно справляется с задачей прогнозирования исполнения государственных контрактов и может быть использована для практических целей. Модель случайного леса также показала неплохие результаты с $R\text{-squared}=0.78$, $MAE=0.18$ и $RMSE=0.24$. Линейная регрессия и дерево решений продемонстрировали более низкие показатели качества. Стоит отметить, что для достижения высокой точности прогнозирования важную роль сыграл тщательный отбор и инженерия признаков, включающих в себя характеристики заказчиков, подрядчиков, самих контрактов и внешние факторы. Это подчеркивает важность этапа подготовки данных при решении задач машинного обучения на практике.

Заключение

В данной работе была рассмотрена возможность применения моделей машинного обучения для прогнозирования исполнения государственных контрактов. Это актуальная задача, так как неисполнение или ненадлежащее исполнение государственных контрактов может приводить к значительным финансовым и организационным потерям. В ходе исследования был проведен анализ характеристик государственных контрактов, влияющих на их исполнение. На основе этого был сформирован набор данных, включающий информацию о заключенных контрактах, характеристики заказчиков и подрядчиков, а также внешние факторы. Данные были тщательно подготовлены для применения моделей машинного обучения. Были протестированы несколько популярных алгоритмов, в том числе

линейная регрессия, деревья решений, случайный лес и XGBoost. Результаты показали, что модель XGBoost продемонстрировала наилучшие показатели качества прогнозирования, с высокой точностью предсказывая вероятность исполнения контрактов. Практическое применение разработанной модели XGBoost может принести значительные выгоды. Она позволит органам государственной власти более эффективно управлять государственными закупками, выявляя контракты, которые с высокой вероятностью могут быть не выполнены вовремя или с нарушениями. Это даст возможность принимать своевременные меры для предотвращения срывов сроков и бюджетов исполнения контрактов.

В дальнейшем исследования в этом направлении могут быть продолжены, в частности, за счет расширения набора данных, добавления новых признаков, а также совершенствования моделей машинного обучения. Комплексное применение технологий прогнозной аналитики и компьютерного зрения в сфере государственных закупок способно значительно повысить эффективность и прозрачность управления государственными контрактами.

Литература

1. Горохова Д. В. Государственные закупки в Российской Федерации: ретроспектива и развитие // Финансовый журнал. – 2020. – Т. 12. – №. 2. – С. 57-68.
2. 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. URL: ivdon.ru/ru/magazine/archive/n8y2023/8600.

3. Борисова В. В. Экосистема государственных закупок //Известия Санкт-Петербургского государственного экономического университета. – 2020. – №. 2 (122). – С. 86-91.

4. Легчаев Р. А. Государственные закупки в условиях цифровизации экономики //Экономика и бизнес: теория и практика. – 2021. – №. 3-2. – С. 49-52.

5. Федорова И. Ю., Пипия Ю. С. Организация системы государственных закупок в Японии и возможности опыта их применения в России //МИР (Модернизация. Инновации. Развитие). – 2021. – Т. 12. – №. 1. – С. 34-55.

6. Мелентьев В.В., Сердечный Д.В., Никитин П.В., Корчагин С.А. Разработка парсинговой системы для анализа государственных контрактов // Информационно-измерительные и управляющие системы. – 2023. – Т. 21. – № 5. – С. 36-47.

7. Гаврилов В.С., Корчагин С.А. Разработка математической модели и программного комплекса для автоматизации научных исследований в области анализа новостей финансовой отрасли // Инженерный вестник Дона, 2024, № 2. URL: ivdon.ru/ru/magazine/archive/n2y2024/9034.

8. Михеева А. А. Применение технологий искусственного интеллекта в сфере государственных закупок //Интерэкспо Гео-Сибирь. – 2022. – Т. 7. – №. 2. – С. 126-130.

9. Максименко М. Р., Максименко Д. Д. Использование открытых данных государственных закупок для анализа рынков лекарственных препаратов //Редакционная коллегия. – 2023. – С. 128.,

10. Кашин Д. В., Шадрин Е. В., Виноградов Д. В. Государственные приоритеты в закупках: результаты опроса российских заказчиков //Наука Юга России: достижения и перспективы. – 2022. – С. 149.



References

1. Gorokhova D. V. Finansovyy zhurnal. 2020. T. 12. №. 2. pp. 57-68.
2. Korchagin S.A., Dogadina E.P., Melentyev V.V., Nikitin P.V. Serdechnyj D.V. Inzhenernyj vestnik Dona. 2023. № 8. URL: ivdon.ru/ru/magazine/archive/n8y2023/8600.
3. Borisova V. V. Izvestiya Sankt-Peterburgskogo gosudarstvennogo ekonomicheskogo universiteta. 2020. №. 2 (122). pp. 86-91.
4. Legchayev R. A. Ekonomika i biznes: teoriya i praktika. 2021. №. 3-2. pp. 49-52.
5. Fedorova I. Yu, Pipiya Yu. S. MIR (Modernizatsiya. Innovatsii. Razvitiye). 2021. vol. 12. №. 1. pp. 34-55.
6. Melentyev V.V., Serdechnyj D.V., Nikitin P.V., Korchagin S.A. Informatsionno-izmeritelnyye i upravlyayushchiye sistemy. 2023. vol. 21. № 5. pp. 36-47.
7. Gavrilov V.S., Korchagin S.A. Inzhenernyj vestnik Dona. 2024, № 2. URL: ivdon.ru/ru/magazine/archive/n2y2024/9034.
8. Mikheyeva A. A. Interekspo Geo-Sibir. 2022. vol. 7. №.2. pp. 126-130.
9. Maksimenko M. R., Maksimenko D. D. Redaktsionnaya kollegiya. 2023. P. 128.
10. Kashin D. V., Shadrina E. V., Vinogradov D. V. Nauka Yuga Rossii: dostizheniya i perspektivy. 2022. P. 149.

Дата поступления: 4.08.2024

Дата публикации: 13.09.2024