
ОТРАСЛЕВЫЕ ПРОБЛЕМЫ

**Прогнозирование вероятности незакрытия торгов
по государственной закупке лекарственных препаратов**

© 2024 г. А.И. Денисова, Д.А. Созаева, К.В. Гончар, Г.А. Александров

А.И. Денисова,

Государственный университет управления, Москва; e-mail: a.i.denisova@inbox.ru

Д.А. Созаева,

Государственный университет управления, Москва;

e-mail: dasozaeva@gmail.com

К.В. Гончар,

Государственный университет управления, Москва; e-mail: goncharkv@gmail.com

Г.А. Александров,

УЦ «ПРОГОСЗАКАЗ.РФ», Омск; e-mail: grishaalex@gmail.com

Поступила в редакцию 15.02.2024

Исследование выполнено при финансовой поддержке Российского научного фонда (проект 23-28-01644).

Аннотация. Качество и своевременность лекарственного снабжения системы здравоохранения посредством проведения государственных закупок является актуальной задачей государственной политики во всех странах мира, включая Россию. Незакрытие (срыв торгов, расторжение уже заключенных контрактов) закупочных процедур в такой социально значимой сфере несет риски для населения, провоцирует возникновение скрытых транзакционных затрат для бюджетной системы, связанных с устранением последствий неудач в закупках. Авторский коллектив в предыдущих работах выявил факторы, которые приводят к незакрытию торгов по закупке лекарственных препаратов и первично оценил последствия их влияния на процедуры закупок. Цель же данной статьи — отталкиваясь от полученных результатов, представить математическую модель вероятности незакрытия торгов. Для достижения поставленной цели путем обработки более 1 млн извещений о государственных закупках лекарственных препаратов за 2022–2023 гг., собранных из открытых источников, выполнены задачи, методологически не решенные до настоящего времени. Так, составлен набор признаков, сопутствующих незакрытию процедур по закупке лекарственных препаратов; проанализирован состав выделенных признаков, оценено их влияние на факт незакрытия процедуры; построена модель вероятности незакрытия процедуры; проанализированы результаты. В отличие от ранее опубликованных исследований прогнозная модель реализована на ансамблях деревьев решений способом градиентного бустинга. Это позволило существенно повысить качество прогноза по каждому фактору, влияющему на вероятность незакрытия торгов. Полученные в статье результаты обладают не только научной новизной, но и могут быть использованы органами регулирования и контроля в сфере государственных закупок для выработки методических рекомендаций заказчикам, направленных на установление оптимальных условий заключения и исполнения контрактов, что приведет к снижению рисков закупок и ущерба для государственного бюджета.

Ключевые слова: государственные закупки лекарственных препаратов, несостоявшиеся процедуры, вероятность незакрытия торгов по государственным закупкам, модель на градиентном бустинге, риски государственных закупок.

Классификация JEL: C55, C57, F12.

УДК: 338.27.

Для цитирования: Денисова А.И., Созаева Д.А., Гончар К.В., Александров Г.А. (2024). Прогнозирование вероятности незакрытия торгов по государственной закупке лекарственных препаратов // *Экономика и математические методы*. Т. 60. № 4. С. 65–76. DOI: 10.31857/S0424738824040066

ВВЕДЕНИЕ

Стремительное развитие информационных технологий и решений в последние несколько лет существенно изменило качество экономических отношений между агентами рынка. В системе государственных закупок также повысились скорость проведения конкурентных процедур и контрактации, практически полностью были оцифрованы процессы исполнения контрактов. Вместе с тем нагрузка на специалистов закупочных служб по-прежнему высока: это — и необходимость тщательно проверять контрагентов, и обязанность корректно подготавливать документацию по закупке, и, конечно же, определять цены контракта. В таких социально значимых и важных отраслях, как здравоохранение, образование, ЖКХ и т.д., ошибки заказчиков или стечение обстоятельств, приводящие к затягиванию или срыву сроков проведения закупок, заключения, исполнения контрактов, являются критическими. Отдельно стоит выделить систему здравоохранения. Здесь, наряду с тем что закупщику необходимо разбираться вкупаемых лекарственных препаратах, от него требуется еще и обоснование цены с обязательным применением сразу нескольких методов расчета, среди которых выбирается наименьший результат.

В отличие от других сфер, где расчет цены контракта проводится одним методом, требуется в 2–3 раза больше времени, вероятность ошибки, соответственно, также возрастает (Денисова и др., 2023, 2024). Неудачная закупка приводит к необходимости повторить весь цикл подготовительных работ, что ведет к скрытым транзакционным затратам для бюджетной системы. Увеличивается риск установления заказчиками непривлекательных цен и других условий исполнения контрактов, что снижает привлекательность закупок в глазах поставщика (подрядчика, исполнителя). В этой связи актуальным становится применение современных вычислительных технологий и решений для предсказания последствий действий, предпринимаемых закупщиком, в частности, оценка влияния установленных параметров торгов при их закрытии (привлекательность для участников). Учитывая значительные объемы государственных закупок в общем объеме ВВП, а также востребованность лекарственных средств, необходимо развить методологии и инструментарий для прогнозирования и моделирования вероятности незакрытия¹ торгов с оценкой ключевых факторов риска провала закупочной процедуры. В этой связи цель нашего исследования — построить модель вероятности незакрытия торгов для государственной закупки лекарственных препаратов в рамках контрактной системы (Закон № 44-ФЗ) под влиянием гетерогенных факторов. Объект исследования — конкурентные процедуры при государственной закупке лекарственных препаратов. Предмет исследования — вероятность незакрытия процедур при государственной закупке лекарственных препаратов и факторы, на нее влияющие.

ОБЗОР ЛИТЕРАТУРЫ И ИССЛЕДОВАНИЙ

Необходимость моделирования и прогнозирования вероятности незакрытия торгов для государственной закупки лекарственных препаратов обосновывается важностью эффективного управления закупками в области здравоохранения. Сфера государственных закупок лекарственных средств имеет критическое значение для общественного здоровья, а вопросы качественного снабжения лекарствами учреждений здравоохранения, отбора добросовестных контрагентов регулярно исследуются зарубежными и российскими учеными (Федоренко, Михайлова, 2015; Hamzah, Perera, Rannan-Eliya, 2020).

Данное исследование также является продолжением серии работ авторов, посвященных оценке эффективности государственных закупок лекарственных препаратов, информационно-аналитическим инструментам сопровождения торгов (Денисова и др., 2023, 2024) и развивает обозначенные в них расчетные и методические вопросы, связанные с использованием продвинутого математического и алгоритмического инструментария.

Проблема оценки и моделирования вероятности наступления какого-либо события стали предметом многих исследований, особенно в контексте управления рисками. В качестве примеров можно привести следующие: предсказание совершения противоправного действия или мошенничества в различных сферах (Денисова, 2017); наступление дефолта инвестиционного проекта и иные события, которые могут стать источниками проектных рисков (Моргунов, 2016); срыва контрактации;

¹ Для краткости будем считать те «незакрытыми процедурами», которые были признаны несостоявшимися в рамках ст. 52 Федерального закона от 05.04.2013 г. № 44-ФЗ «О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд» и в рамках которых не было контрактации.

аварии (например, в сфере энергетики (Sulaeman et al., 2017)); изменение уровня цен, цепочки поставок; предрасположенности пациента к определенному заболеванию (Wang et al., 2021) и т.д. Для решения этих задач в основном применяются различные методы и модели бинарной классификации, суть которых — определение принадлежности наблюдения к одному из двух классов: «1» (событие произойдет) или «0» (событие не произойдет). Кроме того, бинарная классификация — одна из задач машинного обучения «с учителем». Для решения такой задачи могут быть применены такие методы, как модели бинарного выбора (логит-регрессия, пробит-регрессия и др.), деревья случайных решений (random decision tree), машины опорных векторов (support vector machine, SVM) и др., а также методы глубокого обучения (нейронные сети) (Brink, Richards, Fetherolf, 2016).

Безусловно, в связи с необходимостью учета большого объема информации и сложных взаимосвязей между множеством разнородных факторов применение методов анализа больших данных и специальных техник и алгоритмов машинного обучения представляется не только целесообразным, но и обоснованным. Сегодня одним из наиболее результативных методов признается построение ансамблей деревьев случайных решений способом градиентного бустинга (gradient boosting on decision trees, GBDT). Алгоритм заключается в построении множества деревьев, причем каждое следующее строится на остатках (ошибках, отклонениях) предыдущего, причем в качестве минимизации функции ошибок используется метод градиентного спуска. Градиентный бустинг строится на ансамбле моделей со слабой предсказательной силой, чаще всего — именно на деревьях случайных решений (Natekin, Knoll, 2013). Многие исследователи (Dyer et al., 2022; Jun, 2021; Гришин, Строев, 2020; Sudakov, Burnaev, Koroteev, 2019) и практики² признают, что в различных задачах, где используются неоднородные данные, алгоритмы градиентного бустинга и нейронных сетей на сегодняшний день считаются одними из основных подходов. Они позволяют эффективно решать задачу классификации, причем значимых преимуществ алгоритмы по сравнению друг с другом либо не имеют, либо лучшим с точки зрения анализа и простоты разработки признается градиентный бустинг (в то же время нейронные сети показывают лучшие результаты в задачах распознавания изображений).

В контексте государственных закупок применение информационно-аналитического инструментария, и в частности алгоритмов бинарной классификации, обычно обусловлено влиянием рисков — например коррупции или мошенничества (Blafák, 2016; Fazekas, Tóth, King, 2016; Velasco et al., 2021; Sales, 2013), и устойчивыми стратегиями агентов рынка закупок (Лычков, Созаева, Гончар, 2023). Вместе с тем до конца неисследованными остаются проблемы, работа над которыми с помощью методов анализа больших данных позволит повысить эффективность и качество закупочной деятельности, что особенно значимо в такой сфере, как закупка лекарственных препаратов. В отличие от ранее опубликованных исследований, в том числе и авторов статьи (Денисова и др., 2024), подход, предлагаемый в данной работе, позволит повысить качество прогнозирования, проанализировать влияние различных факторов, снизить скрытые затраты на повторное проведение закупок, снизить риски расторжения контрактов, так как *закупки изначально будут объявляться на выгодных условиях*. Это и определило научную новизну исследования и его практическую значимость.

МАТЕРИАЛЫ И МЕТОДЫ

В качестве исходной информационной базы были использованы сведения о проведенных закупках лекарственных препаратов по Закону № 44-ФЗ с извещениями³, опубликованными с января 2022 по октябрь 2023 г. В получившейся таблице — 1 084 921 строка с детализацией товарных позиций: если в процедуре закупается n позиций, то информация о процедуре занимает n строк.

Всего рассматривается 245 006 процедур для закупки лекарственных препаратов, причем в состав товарных позиций этих закупок никакие прочие товары и услуги не входят. Был сформирован набор категориальных, количественных и бинарных факторов, приводящих к незакрытию торгов и отсутствию заявок по ним, которые можно получить исходя из документации (или извещения). На этих факторах была построена модель вероятности незакрытия процедуры по государственной закупке лекарственных препаратов.

² Учебник по машинному обучению. 2.5. Градиентный бустинг (<https://education.yandex.ru/handbook/ml/article/gradientnyj-busting/>).

³ Так называется объявление о закупке в № 44-ФЗ (https://www.consultant.ru/document/cons_doc_LAW_144624/57e264699075c4dccb33283209e1d93f4bc1a3d1/).

В качестве метода моделирования был использован градиентный бустинг над деревьями решений, реализованный с помощью библиотеки `catboost`⁴ на языке Python, которая отличается высокой точностью и возможностью эффективно работать не только с количественными, но и категориальными факторами (Dorogush, Ershov, Gulin, 2018; Bentéjac, Csörgo, Martínez-Muñoz, 2019).

На построенной модели исследуется важность *каждого признака* (feature importance) (Adler, Painsky, 2023), которая показывает, на сколько в среднем изменится результат при изменении значения каждого признака в отдельности. Значение *каждого признака* (feature importance) нормализуется так, чтобы их сумма для всех признаков равнялась 100 (значения *каждого признака* — всегда неотрицательные)⁵.

Для более глубокого анализа и интерпретации влияния наиболее значимых признаков на результат модели были использованы значения Шепли (здесь — SHapley Additive exPlanations, SHAP-значения), полученные с помощью библиотек `catboost` и `shap`⁶. Эти значения прежде всего использовались в теории игр, но были затем адаптированы к моделям машинного обучения: в качестве вклада игроков рассматривается влияние каждого фактора на результат (Lundberg, Erion, Lee, 2018).

ОСНОВНАЯ ЧАСТЬ ИССЛЕДОВАНИЯ

Для достижения цели были поставлены следующие задачи:

- 1) составить набор признаков, сопутствующих незакрытию процедуры по закупке лекарственных препаратов; получить выборку данных по исследуемому периоду;
- 2) проанализировать состав выделенных признаков и оценить их влияние на целевую переменную (факт незакрытия процедуры по причине отсутствия поданных заявок и, соответственно, — гарантированном отсутствии контрактиции);
- 3) построить модель вероятности незакрытия процедуры;
- 4) проанализировать результаты и оценить важность факторов.

Данное исследование логически продолжает и развивает предыдущие работы авторов в области эмпирического анализа факторов со стороны заказчика и поставщика, влияющих на вероятность незакрытия процедур закупки лекарственных препаратов. Ранее было установлено, что «факторы заказчика» (когда заказчик установил невыгодные условия закупки) являются главенствующими в списке случаев незакрытия торгов.

Последствия влияния «факторов заказчика» бывают разными, но самый распространенный случай — когда условия, установленные заказчиком, столь невыгодны, что участники вообще не подают заявок.

Рассмотрим ситуацию полного отсутствия интереса к закупке подробнее.

Для этого определим целевую переменную: значение «1» присваивается, если процедура признана несостоявшейся согласно ст. 52 Закона № 44-ФЗ (что зафиксировано в опубликованном протоколе) и на процедуру не было подано ни одной заявки от поставщиков; в противном случае присваивается «0».

Рассмотрим состав факторов, влияющих на вероятность незакрытия закупки, которые мы можем оценить.

В целом на вероятность незакрытия могут влиять:

- 1) общие факторы (которые могут характеризовать процедуры закупки любых товаров и услуг, например, форма проведения торгов, регион заказчика и т.д.);
- 2) факторы времени (сезон (или месяц) публикации процедуры, время сбора заявок, день недели публикации процедуры и т.д.);
- 3) факторы, связанные с закупаемыми товарами (состав, объем, цена и т.д.).

⁴ CatBoost (<https://catboost.ai/en/docs/>).

⁵ CatBoost. Feature importance (<https://catboost.ai/en/docs/concepts/fstr>).

⁶ SHAP documentation (<https://shap.readthedocs.io/en/latest/index.html>).

При составлении перечня факторов первой группы, авторы прежде всего учитывали ту информацию, которую заказчик обязан предоставлять при публикации извещения. К этим факторам были отнесены:

- электронная торговая площадка (далее — ЭТП) (категориальная переменная);
- форма проведения процедуры (аукцион, запрос котировок и т.п.) (категориальная);
- заказчик (категориальная);
- регион заказчика (категориальная);
- начальная (максимальная) цена контракта НМЦК (количественная);
- является ли закупка совместной (бинарная — может принимать только значения «1» или «0»);
- относительный и абсолютный размер обеспечения заявки (количественная);
- относительный и абсолютный размер аванса (количественная);
- осуществляется ли закупка только у субъектов малого предпринимательства (бинарная);
- является ли закупка централизованной и предназначенной субъектам малого предпринимательства (бинарная);
- число поставщиков в регионе заказчика, хотя бы один раз за прошедший год осуществивших поставку лекарственных препаратов (количественная).

Далее были выделенные новые признаки второй группы, описывающие временные характеристики процедуры:

- продолжительность сбора заявок в днях (количественная);
- месяц публикации извещения о процедуре (категориальная);
- день недели начала приема заявок (категориальная);
- день недели окончания приема заявок (категориальная).

Более подробно остановимся на формировании факторов третьей группы. Существуют признаки, которые описывают отдельный лекарственный препарат:

- фигурирует ли лекарственный препарат в списке жизненно необходимых и важнейших лекарственных препаратов (далее — ЖНВЛП) (бинарная);
- присутствуют ли в составе препарата наркотические вещества (бинарная);
- является ли препарат дефектурным⁷ (согласно итогам 2022 г.)⁸ (бинарная);
- существует ли в рамках международного непатентованного наименования закупаемого лекарственного препарата не более двух торговых наименований (бинарная);
- число торговых наименований лекарственного препарата (количественная);
- число торговых наименований лекарственного препарата, у которых страной производства является Российская Федерация (количественная);
- доля торговых наименований лекарственного препарата, у которых страной производства является Российская Федерация (количественная);
- фармакотерапевтическая группа (далее — ФТГ) препарата (категориальная);
- наименование класса ОКПД (8 цифр), к которому принадлежит лекарственный препарат (категориальная);
- наименование лекарственного препарата по справочнику ЕСКЛП⁹ (категориальная);
- число поставщиков в регионе заказчика, хотя бы один раз за прошедший год осуществивших поставку лекарственных препаратов определенных наименований.

⁷ Отсутствует в аптеках, является узкоспециализированным, отсутствует у поставщиков в связи с перерегистрацией, браком серий или проблемами с ввозом препарата в страну (*Прим. ред.*).

⁸ «Правительство выявило потенциально дефектурные лекарства» // *Фармацевтический вестник*. 25.07.2022 (<https://pharmvestnik.ru/content/news/Pravitelstvo-vyyavilo-potencialno-defekturnye-lekarstva.html>).

⁹ Единый структурированный справочник-каталог лекарственных препаратов (<https://esklp.egisz.rosminzdrav.ru/>).

Заметим здесь, что объектом наблюдения в рамках построения модели является *процедура*, а не лекарственный препарат. Соответственно, необходимо применить технику агрегирования данных об отдельных препаратах так, чтобы составить описание процедуры, в рамках которой они закупаются. Был применен следующий подход:

- для агрегирования бинарных факторов применялось логическое сложение (например, если хотя бы один препарат из закупки входит в перечень ЖНВЛП (для хотя бы одного препарата значение «1», то для всей процедуры в целом тоже будет стоять значение «1»));
- для агрегирования количественных факторов использовалось минимальное значение;
- для агрегирования категориальных факторов использовался метод «one-hot encoding», т.е. кодирование каждого значения исходного фактора новой бинарной переменной, которая показывает принадлежность к этому уникальному значению. Например, для признака ФТГ это будет интерпретироваться как «наличие в закупке хотя бы одного препарата, относящегося к соответствующей фармакотерапевтической группе».

Для закупаемых лекарственных препаратов в большинстве случаев также известны требуемый объем и цена единицы объема. При этом единицы измерения отличаются в зависимости от формы и действующего вещества лекарства, т.е. напрямую агрегировать этот показатель нельзя.

Проведем нормирование объемов закупаемых препаратов в рамках потребительских единиц лекарственных препаратов. Основных наименований форм выпуска можно выделить семь — они встречаются в 96% наблюдений. Это: «г (граммы) действующего вещества», «г лекарственной формы», «доз(а)», «капсула», «мг действующего вещества», «мл» и «таблетка».

Приведем к интервалу от 0 до 1 объемы разных препаратов в рамках каждой единицы измерения.

Для основных выделенных единиц нормирование осуществлялось по интервалам следующим образом:

- все значения закупаемых объемов в рамках каждой единицы измерения были разделены на определенное число интервалов одинаковой длины;
- каждому интервалу присваивалось значение от 0 до 1 так, чтобы 0 соответствовал нулевому объему (это означает, что у заказчика не было возможности описать требуемый объем лекарственного препарата), 1 — интервалу, в котором содержатся наибольшие значения. Всем остальным интервалам были присвоены значения нарастающим итогом с шагом, равным отношению 1 к числу интервалов.

Для получения нормированного объема $q_{norm}(i)$ лекарственного препарата i при прочих единицах измерения использовали формулу:

$$q_{norm}(i) = (q(i) - q_{min}) / (q_{max} - q_{min}), \quad (1)$$

где q_{min} — минимальное значение среди закупаемых объемов лекарственных препаратов при прочих единицах измерения; q_{max} — максимальное значение среди закупаемых объемов лекарственных препаратов при прочих единицах измерения.

В результате для каждой процедуры мы получили значение нового признака — суммы нормированного объема закупаемых лекарственных препаратов.

Интересным также является сравнение объявленных цен на лекарственные препараты и их средневзвешенных цен по регионам. Создадим в связи с этим следующие новые факторы:

- число лекарственных препаратов в закупке, у которых цена меньше средневзвешенной по региону за исследуемый период (количественная);
- доля лекарственных препаратов в закупке, у которых цена меньше средневзвешенной по региону за исследуемый период (количественная);
- минимальное отклонение цен препаратов в закупке от соответствующих средневзвешенных цен по региону за исследуемый период (количественная);
- максимальное отклонение цен препаратов в закупке от соответствующих средневзвешенных цен по региону за исследуемый период (количественная).

Максимальное и минимальное отклонения могут принимать как положительные, так и отрицательные значения.

Объявленные цены на препараты не являются в полной мере «чистыми» (т.е. это — цены на упакованный определенным образом товар), поскольку зависят и от требуемых объемов, и от способа фасовки. Для того чтобы проанализировать препараты по цене, построим их ранжир по стоимости.

Выполним следующую процедуру.

1. Для каждого препарата в рамках всех регионов были оценены среднерыночные цены на основе всех опубликованных закупок за исследуемый период.

2. По получившейся выборке цен определили аномальные значения — те, которые не входят в 95% выборки. Получившиеся аномальные значения представляют собой множество элементов с одинаковым рангом по цене.

3. Исключаем аномальные наблюдения из выборки. Оцениваем коэффициент вариации и снова определяем аномальные значения. Новая группа выбросов представляет собой множество элементов со следующим рангом.

4. Продолжаем до тех пор, пока значение коэффициента вариации практически не перестанет уменьшаться при исключении выбросов из выборки.

В итоге у нас получилось 17 групп препаратов. Множеству препаратов с наименьшими ценами присвоен ранг 0; множеству с наибольшими — ранг 16.

В результате для каждой процедуры получим значения новых признаков:

- минимальный ранг по стоимости среди препаратов в закупке (категориальная);
- максимальный ранг по стоимости среди препаратов в закупке (категориальная).

Следующий шаг заключается в итоговой агрегации данных и формировании обучающей и тестовой выборок. Использовать в модели одновременно и код ОКПД, и ФТГ, и наименование препарата, согласно ЕСКЛП, может быть нецелесообразно, так как это значительно увеличит размерность данных, но незначимо скажется на качестве модели, поскольку все эти факторы разной степени описывают применимость и свойства препаратов.

После применения процедуры кодирования (one-hot encoding) на основе 95 уникальных значений ОКПД и с учетом остальных факторов получился набор из 125 признаков, описывающих все доступные базовые, временные, стоимостные и товарные характеристики процедуры.

Обучающая выборка составляет 80% всех процедур, 20% — тестовая. Наиболее подходящие значения параметров модели, такие как число деревьев, их глубина, скорость обучения, были подобраны с помощью алгоритма поиска по сетке (grid search).

Точность (ассигасу), т.е. доля верно классифицированных исходов на тестовой выборке, построенной модели градиентного бустинга с помощью catboost составляет 84,0% (на обучающей — 90,0%). Чувствительность модели (доля верно классифицированных незакрытых процедур) составляет 83,98%, специфичность (доля верно классифицированных успешно закрытых процедур) — 83,60%, что оценивается как достаточно высокие результаты.

Согласно расчетам, представленным на рис. 1, наибольшую важность (feature importance) и влияние на результат имеют следующие факторы:

- НМЦК (lot_price_correct), максимальное (max_share_price) и минимальное (min_share_price) отклонение цен препаратов в закупке от соответствующих средневзвешенных цен по региону за исследуемый период;
- заказчик (cust); минимальная доля отечественных торговых наименований среди препаратов в закупке (share_rus_name);
- регион заказчика (cust_reg);
- минимальное число отечественных торговых наименований среди препаратов в закупке (count_rus_name);
- минимальное число поставщиков, которые ранее поставляли препараты в закупке в данный регион (count_sup_by_ftg);
- минимальное число торговых наименований среди препаратов в закупке (count_trade_name);

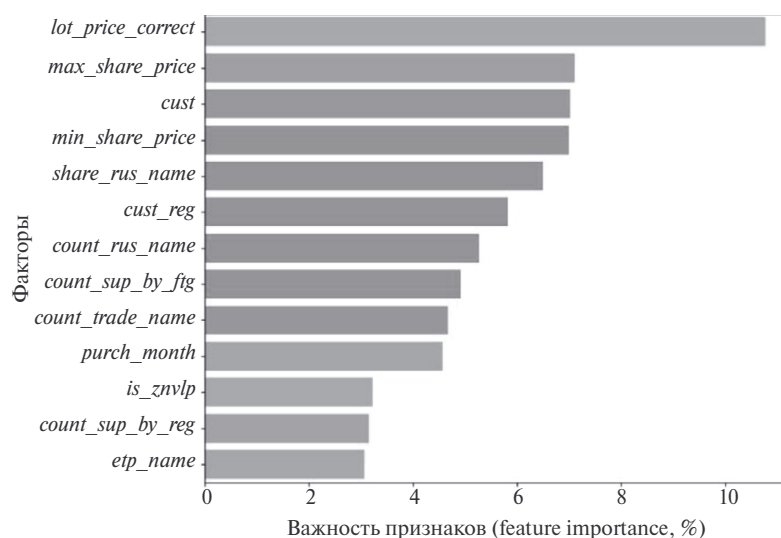


Рис. 1. Наиболее значимые факторы, влияющие на незакрытие торгов

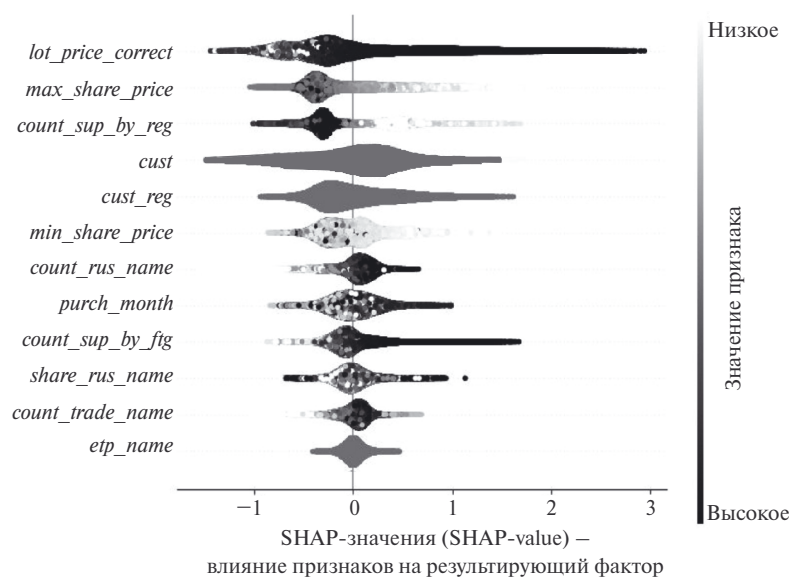


Рис. 2. SHAP-значения наиболее значимых факторов незакрытия торгов¹¹

- месяц публикации процедуры (purch_month);
- наличие в закупке хотя бы одного препарата из списка ЖНВЛП (is_znvlp);
- минимальное число поставщиков лекарственных препаратов в данном регионе (count_sup_by_reg);
- ЭТП (etp_name).

Сумма взвешенных значений feature importance по этим факторам превышает 70%, все прочие признаки имеют меньшую значимость (рис. 1).

На рис. 2 представлена диаграмма SHAP-значений наиболее важных факторов модели¹⁰. Каждая процедура в обучающей выборке представляется в виде точки, которая окрашивается

¹⁰ Полноцветные рисунки доступны по запросу у авторов.

¹¹ cust, cust_reg, etp_name — категориальные переменные.

в соответствии со значением (feature value) исследуемого признака. Расположение и цвета множества точек позволяют судить о распределении влияния каждого фактора на результат.

Разработанный выше машинный алгоритм лег в основу прототипа, прогностического (т.е. наш ресурс — как прогнозная матрица, которая выдает пользователю смоделированные последствия его управленческих решений) ИТ-решения, разработанного авторами для поддержки управленческих решений специалистов закупочных служб медицинских учреждений.

Описанная модель была реализована в приложениях¹² BI Qlik Sense. Пользователь (заказчик) должен ввести информацию о закупке, которую он собирается провести, заполнить список закупаемых позиций. После этого будет рассчитана вероятность незакрытия процедуры.

Поскольку применение технологий машинного обучения специалистами в сфере закупок требует специальных навыков и компетенций, что не предусмотрено их профессиональным стандартом, то самостоятельно пройти по методологии, предложенной авторами, они просто не смогут. В этой связи удобная визуализация в приложении позволит заказчикам наглядно видеть узкие места своей закупочной процедуры и оперативно их скорректировать.

РЕЗУЛЬТАТЫ И ВЫВОДЫ

Из данных рис. 2 следует, что высокие значения НМЦК снижают вероятность незакрытия. Однозначной рекомендацией заказчику является корректное определение цен на лекарственные препараты путем сравнения их со средней по рынку, при этом, чем выше цена препарата, тем ниже вероятность незакрытия.

Наличие среди закупаемых препаратов более дорогих или более дешевых (по признакам максимального и минимального ранга препаратов по ценам) на результат влияет гораздо меньше. Тем не менее процедуры закупки дорогих лекарственных препаратов (даже в сочетании с дешевыми) закрываются немного охотнее.

По итогам расчетов можно сделать вывод о том, что не так важны характеристики закупаемых позиций (в особенности закупаемые объемы), как их суммарная стоимость — НМЦК (см. рис. 1).

Факторы времени в целом менее значимы: наиболее серьезно на вероятность незакрытия влияет *месяц публикации*; на втором месте — длительность приема заявок.

Такие факторы, как обеспечение заявки, аванс, совместное проведение закупки, проведение закупки для субъектов малого предпринимательства, не влияют значимо на результат.

Из рис. 2 также можно заметить, что:

- увеличение числа отечественных торговых наименований у лекарственных препаратов, а также, в меньшей степени, общего числа торговых наименований ведет к уменьшению вероятности незакрытия;

- при этом большая часть наблюдений имеет такие значения числа торговых наименований, которые приводят скорее к повышению вероятности незакрытия;

- медицинские учреждения, как правило, закупают препараты именно из списка ЖНВЛП, как следует из содержания ст. 80 Федерального закона от 21.11.2011 № 323-ФЗ (в ред. от 28.12.2022) «Об основах охраны здоровья граждан в Российской Федерации», чем и обусловлены такие результаты;

- вероятность незакрытия процедур, опубликованных во второй половине года, незначительно ниже, чем в первой¹³;

- на вероятность незакрытия процедуры сильнее влияет наличие достаточного числа специализированных поставщиков в регионе (т.е. тех, которые ранее поставляли конкретный препарат), чем тех, которые просто поставляли лекарства;

¹² Работа, направленная на создание приложения, осуществлялась на базе федерального оператора торгов АО «ЕЭТП» («Росэлторг»).

¹³ Активность проведения торгов характеризуется определенными сезонными колебаниями, пики активности которых, как правило, приходится на весну и на конец года. Поскольку чаще всего не подразумевается, что спрос заказчика может быть отложен, указание конкретных месяцев в качестве рекомендаций не имеет смысла.

— увеличение числа специализированных поставщиков в регионе ведет к уменьшению вероятности незакрытия процедуры.

Рассмотрим теперь факторы «заказчик», «регион заказчика» и «наименование электронной торговой площадки (ЭТП)». Значимого смещения числа этих объектов в сторону более удачного завершения процедур по графику на рис. 2 не наблюдается. Регион местонахождения заказчика, сама фигура заказчика, как и ЭТП, не являются, по нашему мнению, факторами, которые могут существенно влиять на вероятность незакрытия процедур. Это — объективные факторы: есть регионы, где объем закупок лекарств больше, и, соответственно, на их территории находится больше заказчиков или они более крупные. В таком случае и статистика незакрытия процедур здесь будет выше. Сменить регион регистрации заказчика, даже если имеются элементы регионального протекционизма, местные национальные, климатические особенности, не представляется возможным. Аналогичная ситуация и с ЭТП: число незакрытых процедур напрямую зависит от общего объема проводимых на площадке торгов.

ЗАКЛЮЧЕНИЕ

Данная статья является завершающей работой цикла статей авторов о масштабном исследовании параметров рынка государственных закупок лекарственных средств.

Благодаря проведенным изысканиям авторы смогли всесторонне исследовать факторы, влияющие на успешность, экономичность, качество контрактации в системе государственных и муниципальных закупок. К наиболее значимым результатам, полученным в рамках данной статьи, можно отнести высокоточное исследование факторов, влияющих на незакрытие торгов по закупке лекарств с применением ранее неиспользованного для этого аналитического инструментария.

Полученные выводы могут быть направлены в профильные министерства и ведомства для совершенствования работы, связанной с лекарственным обеспечением учреждений здравоохранения; в органы регулирования и контроля закупочной деятельности для повышения качества методической работы, выработки рекомендаций заказчикам, направленным на снижение рисков и преодоление проблем, связанных с организацией и проведением закупок.

Дальнейшее развитие исследования авторы видят во внедрении его результатов в практическую деятельность вышеперечисленных организаций и развитие созданного авторами прототипа программного обеспечения. Одним из возможных путей применения полученных результатов может быть разработка полноценного платформенного решения для поддержки процессов государственных закупок лекарственных средств в крупных учреждениях здравоохранения.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Гришин А.А., Строев С.П. (2020). Инструментарий и особенности решения задачи классификации в системах кредитного скоринга // *Continuum. Математика. Информатика. Образование*. № 1 (17). С. 51–59. DOI: 10.24888/2500-1957-2020-17-1-51-59 [Grishin A.A., Stroeve S.P. (2020). Tools and features of solving the classification problem in credit scoring systems. *Continuum. Maths. Informatics. Education*, 1 (17), 51–59. DOI: 10.24888/2500-1957-2020-17-1-51-59 (in Russian).]
- Денисова А.И. (2017). Моделирование рисков лицензируемых товарных рынков российской Федерации // *Научные труды Вольного экономического общества России*. Т. 206. № 4. С. 133–145. [Denisova A.I. (2017). Risk modeling in the licensed commodity markets of the Russian Federation. *Scientific Works of the Free Economic Society of Russia*, 206, 4, 133–145 (in Russian).]
- Денисова А.И., Созаева Д.А., Гончар К.В., Александров Г.А. (2023). Совершенствование методологии оценки экономической эффективности государственных закупок лекарственных средств // *Финансовый журнал*. Т. 15. № 4. С. 63–81. DOI: 10.31107/2075-1990-2023-4-63-81 [Denisova A.I., Sozaeva D.A., Gonchar K.V., Aleksandrov G.A. (2023). Improving the methodology for assessing the economic efficiency of public procurement of medicines. *Financial Journal*, 15 (4), 63–81. DOI: 10.31107/2075-1990-2023-4-63-81 (in Russian).]
- Денисова А.И., Созаева Д.А., Гончар К.В., Александров Г.А. (2024). Анализ факторов незакрытия торгов по госзакупкам лекарственных препаратов // *Проблемы прогнозирования*. № 1 (202). С. 178–191. DOI: 10.47711/0868-6351-203-178-191 [Denisova A.I., Sozaeva D.A., Gonchar K.V., Aleksandrov G.A. (2024). Analysis of factors of failure of public procurement of medicines. *Studies on Russian Economic Development*, 1 (202), 178–191. DOI: 10.47711/0868-6351-203-178-191 (in Russian).]

- Лычков И.И., Гончар К.В., Созаева Д.А. (2023). Консерватизм, мобильность, изоляция: подход к исследованию поведения агентов рынка государственных закупок // *Экономика и математические методы*. Т. 59. № 4. С. 45–57. DOI: 10.31857/S042473880026993-1 [Lychkov I.I., Gonchar K.V., Sozaeva D.A. (2023). Conservatism, mobility, isolation: Approach to study of agents' behavior in the public procurement market. *Economics and Mathematical Methods*, 59, 4, 45–57. DOI: 10.31857/S042473880026993-1 (in Russian).]
- Моргунов А.В. (2016). Моделирование вероятности дефолта инвестиционных проектов // *Корпоративные финансы*. Т. 10. № 1 (37). С. 23–45. [Morgunov A.V. (2016). Modeling the probability of default of the investment projects. *Journal of Corporate Finance Research*, 10, 1 (37), 23–45 (in Russian).]
- Федоренко И.Н., Михайлова В.В. (2015). Контроль и оптимизация как основа повышения эффективности расходования бюджетных средств в системе государственных закупок // *Инновационная наука*. № 12-1. С. 311–314. [Fedorenko I.N., Mikhailova V.V. (2015). Control and optimization as a basis for improving the efficiency of budget spending in the public procurement system. *Innovative Science*, 12–1, 311–314 (in Russian).]
- Adler A.I., Painsky A. (2022). Feature importance in gradient boosting trees with cross-validation feature selection. *Entropy*, 24, 687. DOI: 10.3390/e24050687
- Bentéjac C., Csörgo A., Martínez-Muñoz G. (2019). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54, 1937–1967.
- Blafák J. (2016). Reduction of the risk in public procurement by using design-build as a means for suitable constructing. *Stavební Obzor — Civil Engineering Journal*, 25 (2). DOI: 10.14311/CEJ.2016.02.0007
- Brink H., Richards J., Fetherolf M. (2016). *Real-world machine learning*. 1st ed. N.Y.: Manning. 264 p. [Бринк Х., Ричардс Дж., Февееролф М. (2017). *Машинное обучение*. СПб.: Питер. 336 с. (на русс. языке)].
- Dorogush A.V., Ershov V., Gulin A. (2018). *CatBoost: Gradient boosting with categorical features support*. *ArXiv, abs/1810.11363*
- Dyer A.S., Zaengle D., Nelson J.R., Duran R., Wenzlick M., Wingo P.C. et al. (2022). Applied machine learning model comparison: Predicting offshore platform integrity with gradient boosting algorithms and neural networks. *Marine Structures*, 83. DOI: 10.1016/j.marstruc.2021.103152
- Fazekas M., Tóth I.J., King L.P. (2016). An objective corruption risk index using public procurement data. *The European Journal on Criminal Policy and Research*, 22, 369–397. DOI: 10.1007/s10610-016-9308-z
- Hamzah N., Perera P.N., Rannan-Eliya R.P. (2020). How well does Malaysia achieve value for money in public sector purchasing of medicines? Evidence from medicines procurement prices from 2010 to 2014. *BMC Health Services Research*, 20, 509. DOI: 10.1186/s12913-020-05362-8
- Jun M. (2021). A comparison of a gradient boosting decision tree, random forests, and artificial neural networks to model urban land use changes: The case of the Seoul metropolitan area. *International Journal of Geographical Information Science*, 35, 2149–2167.
- Lundberg S.M., Erion G.G., Lee S. (2018). *Consistent individualized feature attribution for tree ensembles*. *ArXiv: abs/1802.03888*
- Natekin A., Knoll A. (2013). Gradient boosting machines, a tutorial. *Front. Neurorobotics*, 7 (21). DOI: 10.3389/fnbot.2013.00021
- Sales L. (2013). Risk prevention of public procurement in the Brazilian government using credit scoring. 2013 OBEGEF — Observatório de Economia e Gestão de Fraude. *Paper in the International Conference Interdisciplinary Insights on Fraud and Corruption. Working Papers*, 19. 27 p.
- Sudakov O., Burnaev E., Koroteev D. (2019). Driving digital rock towards machine learning: Predicting permeability with gradient boosting and deep neural networks. *Computers & Geosciences*, 127, 91–98. DOI: 10.1016/j.cageo.2019.02.002
- Sulaeman S., Benidris M., Mitra J. Singh C. (2017). A wind farm reliability model considering both wind variability and turbine forced outages. *2017 IEEE Power & Energy Society General Meeting*. Chicago (IL), USA, 1–1. DOI: 10.1109/PESGM.2017.8274250
- Velasco R.B., Carpanese I., Interian R., Paulo Neto O.C.G., Ribeiro C.C. (2021). A decision support system for fraud detection in public procurement. *International Transactions in Operational Research*, 28, 27–47. DOI: 10.1111/itor.12811
- Wang B., Liu Y., Sun X. et al. (2021) Prediction model and assessment of probability of incident hypertension: The rural Chinese cohort study. *Journal of Human Hypertension*, 35, 74–84. DOI: 10.1038/s41371-020-0314-8

Predicting the probability of failure in medicines' public procurement

© 2024 A.I. Denisova, D.A. Sozaeva, K.V. Gonchar, G.A. Aleksandrov

A.I. Denisova,

The State University of Management, Moscow, Russia; e-mail: a.i.denisova@inbox.ru

D.A. Sozaeva,

The State University of Management, Moscow, Russia; e-mail: dasozaeva@gmail.com

K.V. Gonchar,

The State University of Management, Moscow, Russia; e-mail: goncharkv@gmail.com

G.A. Aleksandrov,

"PROGOSZAKAZ.RF", Omsk, Russia; e-mail: grishaalex@gmail.com

Received 15.02.2024

The study was supported by the Russian Science Foundation (project 23-28-01644).

Abstract. The quality and timeliness of medicines' supply to the healthcare system through public procurement is an urgent task of public policy in all countries of the world, including Russia. Failure to close (failure of tenders, termination of already concluded contracts) procurement procedures in such a socially significant area carries risks for the population, provokes the emergence of hidden transaction costs for the budget system to eliminate the consequences of procurement failures. In their previous works, the authors identified the factors that lead to the failure to close tenders for the purchase of remedies, and initially assessed the consequences of their influence on procurement procedures. The purpose of this article is to present a mathematical model of the probability of non-closure tenders based on the obtained results. To achieve this goal, through processing more than 1 million notifications of public procurement of remedies for 2022–2023, collected from the open sources, the tasks with no methodological solutions were completed. Thus, a set of features accompanying the failure to close procedures for the purchase of remedies was compiled. The composition of identified features was analyzed; their influence on non-closure of the procedure was assessed. A model of the probability of non-closure of the procedure was constructed, and the results were interpreted. Unlike previously published studies, the forecast model was implemented on the ensembles of decision trees using gradient boosting. This made possible to significantly improve the quality of the forecast for each factor affecting the probability of non-closure of the bidding. The results obtained in the article are not only scientifically novel, but can also be used by regulatory and control bodies in public procurement to develop methodological recommendations for customers on establishing optimal conditions for concluding and fulfilling contracts, which will reduce procurement risks and damage to the state budget.

Keywords: public procurement of remedies, failed procedures, probability of non-closure of tenders for public procurement, gradient boosting model, public procurement risks.

JEL Classification: C55, C57, F12.

UDC: 338.27.

For reference: **Denisova A.I., Sozaeva D.A., Gonchar K.V., Aleksandrov G.A.** (2024). Predicting the probability of failure in medicines' public procurement. *Economics and Mathematical Methods*, 60, 4, 65–76. DOI: 10.31857/S0424738824040066 (in Russian).