

Вестник Евразийской науки / The Eurasian Scientific Journal <https://esj.today>

2018, №3, Том 10 / 2018, No 3, Vol 10 <https://esj.today/issue-3-2018.html>

URL статьи: <https://esj.today/PDF/24ECVN318.pdf>

Статья поступила в редакцию 14.05.2018; опубликована 04.07.2018

Ссылка для цитирования этой статьи:

Перерва О.Л., Степанов С.Е., Незимова С.С. Использование анализа больших данных для определения факторов эффективности процесса государственных закупок // Вестник Евразийской науки, 2018 №3, <https://esj.today/PDF/24ECVN318.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.

For citation:

Pererva O.L., Stepanov S.E., Nezimova S.S. (2018). Using big data analysis to determine the factors of the effectiveness of the public procurement process. *The Eurasian Scientific Journal*, [online] 3(10). Available at: <https://esj.today/PDF/24ECVN318.pdf> (in Russian)

УДК 338 + 519.86

ГРНТИ 06.75

Перерва Ольга Леонидовна

ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
Калужский филиал, Россия
Заместитель директора
Доктор экономических наук, профессор
E-mail: pol@bmstu-kaluga.ru
РИНЦ: http://elibrary.ru/author_profile.asp?id=406482

Степанов Сергей Евгеньевич

ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
Калужский филиал, Россия
Декан факультета
Кандидат физико-математических наук, доцент
E-mail: stepanov@bmstu-kaluga.ru
РИНЦ: http://elibrary.ru/author_profile.asp?id=9853
Researcher ID: <http://www.researcherid.com/rid/U-9190-2017>
Scopus: <https://www.scopus.com/authid/detail.uri?authorId=8328785000>

Незимова София Сергеевна

ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
Калужский филиал, Россия
Магистрант
E-mail: stepanova_s@adm.kaluga.ru

Использование анализа больших данных для определения факторов эффективности процесса государственных закупок

Аннотация. В статье сделана попытка проанализировать эффективность проведения аукциона с точки зрения организаторов. Анализ статистики проведения аукционов в электронной форме показывает, что лишь около 60 % из них являются состоявшимися, поэтому нередко приходится заявки размещать повторно. Это приводит к дополнительным затратам организаторов, заказчиков, торговых площадок и поставщиков. Выявление факторов,

влияющих на эффективность аукционов, и разработка рекомендаций организаторам, торговой площадке и заказчикам является целью данной работы.

Для проведения анализа были взяты данные об электронных аукционах, проводившихся в Калужской области в 2017 году в соответствии с Федеральным законом № 44-ФЗ. В результате обработки исходных данных исследуемое множество состояло из 3890 записей, содержащей информацию по 13 переменным. Авторами использовались такие методы анализа, как корреляционный анализ, проверка статистических гипотез, логистическая регрессия, деревья решений, случайные леса. Выбор именно этих методов объясняется как целью исследования, так и их эффективностью.

Авторами была установлена зависимость между тем, состоялся аукцион или нет и рассматриваемыми факторами. Используя комбинацию различных моделей, выявлены наиболее важные факторы, влияющие на эффективность проведения аукциона.

По результатам анализа данных были сделаны определённые выводы и даны рекомендации. Например, определён диапазон начальной цены, при которой наиболее вероятно, что аукцион состоится. Авторы выявили группы товаров и услуг с различной эффективностью проведения аукционов, что может служить основой для работы между заказчиками, организаторами и поставщиками товаров и услуг.

Ключевые слова: государственные закупки; электронные аукционы; логистическая регрессия; дерево решений; случайные леса

Введение

Процесс государственных закупок, проводимых с 1 января 2014 года в соответствии с Федеральным законом № 44-ФЗ¹, играет существенную роль во взаимодействии государства, общества и субъектов предпринимательства.

Правовые основания и коллизии, возникающие в процессе реального применения этого закона, отражены в работе [7]. Анализ практики государственных и муниципальных закупок у субъектов малого предпринимательства проведён в статье [8]. Процесс проведения аукционов в электронной форме достаточно полно описан в работе [6]. Типичные ошибки государственных и муниципальных заказчиков при осуществлении электронных аукционов достаточно полно представлены в работе [3].

В данной работе сделана попытка проанализировать эффективность проведения аукциона с точки зрения организаторов. В Калужской области эта функция осуществляется управлением государственных закупок Министерства конкурентной политики. Одной из целей работы является максимально эффективное распоряжение бюджетными средствами с целью получения качественных товаров и услуг, обеспечения конкуренции, стимулирования инноваций, ответственности за результативность обеспечения государственных и муниципальных нужд.

Анализ статистики проведения аукционов в электронной форме показывает, что лишь около 60 % из них являются состоявшимися, поэтому нередко приходится заявки размещать повторно. Это приводит к дополнительным затратам организаторов, заказчиков, торговых площадок и поставщиков. Выявление факторов, влияющих на эффективность аукционов, и

¹ Федеральный закон от 5 апреля 2013 года № 44-ФЗ «О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд».

разработка рекомендаций организаторам, торговой площадке и заказчикам является целью данной работы.

Методы

Для проведения анализа были взяты данные с официального сайта единой информационной системы в сфере закупок². Из представленных данных в качестве обучающего множества были выбраны аукционы, проводившиеся в Калужской области в период с января по октябрь 2017 года в соответствии с 44-ФЗ на площадке Сбербанка РФ. Данные за ноябрь и декабрь 2017 года были использованы в качестве тестового множества для проверки качества построенных моделей.

Обучающее множество представляет собой набор из 3975 записей о различных объявленных аукционах. Каждая запись содержит характеристики по 53 параметрам. Для проведения анализа из всех параметров были оставлены двенадцать: дата публикации, дата окончания срока подачи заявок, коды ОКПД, начальная цена, информация о том, является аукцион совместным, является ли он аукционом для субъектов малого предпринимательства, предоставлялись ли преимущества учреждениям и предприятиям уголовно-исполнительной системы, предоставлялись ли преимущества организациям инвалидов, предоставлялись ли преференции, результат аукциона (состоялся или нет), общее количество поданных заявок и процент снижения первоначальной цены по результатам торгов. Два первых параметра были преобразованы в три, на наш взгляд более информативные: день недели публикации, день недели окончания и число дней на оформление заявки. Коды ОКПД, соответствующие отраслям и представляющие девятизначные наборы, были сокращены до укрупнённых, четырёхзначных, групп. Если в записи содержалось несколько кодов, то оставлялся один, наиболее часто встречающийся или наиболее подходящий по смыслу.

Оставление именно этих параметров для анализа эффективности аукциона обусловлено тем, что они или являются целевыми (результат аукциона, количество поданных заявок, процент снижения), или ими можно управлять.

Из 3975 записей были удалены 85, которые соответствовали отменённым аукционам. В результате обработки исходных данных обучающее множество состояло из 3890 записей, содержащей информацию по 13 параметрам. Два из них, количество поданных заявок и процент снижения, являются количественными характеристиками аукциона и позволяют дать статистически обоснованный ответ на вопрос о необходимости привлечения к аукционам большего числа участников.

Из оставшихся 11 параметров один – результат аукциона – является объясняемой бинарной переменной, остальные – объясняющими. Такие параметры как дни недели и коды ОКПД рассматривались как факторные переменные, начальная цена и число дней на оформление заявки – как количественные, остальные как бинарные переменные.

Для проведения анализа были использованы следующие методы статистического анализа, бизнес-аналитики и машинного обучения:

1. Корреляционный анализ. Он применялся для выявления связи между числом участников аукциона и процентом снижения первоначальной цены.

² <http://zakupki.gov.ru>.

2. Проверка статистических гипотез о равенстве средних значений. Применялась для сравнения эффективности проведения аукционов в зависимости от дня недели размещения заявок и дня недели Окончания приёма заявок.

3. Логистическая регрессия. Уравнение логистической регрессии в общем случае имеет вид

$$p(x_1, x_2, \dots, x_n) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}},$$

где переменную p можно интерпретировать как вероятность того, что выходная переменная примет значение 1. Построение модели логистической регрессии использовалось для выявления важности признаков, направления связи между признаками и результатом аукциона. Выбор логистической регрессии объясняется тем, что выходная переменная – результат аукциона – является бинарной (либо аукцион состоялся, либо нет). В качестве критерия важности признаков использовалось вероятность ошибки первого рода при проверке статистической гипотезы о равенстве нулю соответствующего коэффициента. В используемой при проведении расчётов системе статистической обработки данных R эти вероятности рассчитываются автоматически.

4. Деревья решений. При проведении анализа использовалось дерево классификации и регрессии CART. Регрессионный анализ с помощью данной модели позволил разделить коды закупаемых товаров и услуг на несколько множеств с различной средней эффективностью аукционов и дать соответствующие рекомендации для организаций, размещающих заявки. Использование этого же дерева как классификатора часто позволяет выявить факторы, которые наиболее важны для успешности аукциона.

5. Случайные леса. Этот метод является одним из наиболее эффективных классификаторов и позволяет построить наиболее точные предсказания выходной переменной [1]. Кроме того, он позволяет выявить влияние не только одного-двух важнейших факторов, но и учесть вклад остальных признаков в итоговый результат.

Выбор именно этих методов объясняется как целью исследования, так и их эффективностью. Сравнение некоторых из представленных методов и традиционных эконометрических моделей проведено авторами в статье [10].

Результаты

Из 3890 запланированного аукциона состоялись 2266 (аукцион считается состоявшимся, если число поданных и не отозванных заявок на момент окончания не менее двух), что составляет 58,25 %.

1. Выборочный коэффициент корреляции между числом участников состоявшихся аукционов и процентом снижения первоначальной цены составляет $r_g = 0,4465$. Для проверки гипотезы о значимости выборочного коэффициента корреляции рассчитывается величина

$$T = r_g \sqrt{n-2} / \sqrt{1-r_g^2},$$

которая имеет распределение Стьюдента с $n-2$ степенями свободы и является наблюдаемым значением критерия проверки [4].

Для полученных данных величина $T = 23,75$. Сравнивая её с критической точкой распределения Стьюдента $t_{кр}(0,01;2264) = 2,578$, убеждаемся, что $T > t_{кр}$ и, следовательно, выборочный коэффициент корреляции значимо отличается от нуля, т. е. число участников состоявшихся аукционов и процент снижения первоначальной цены имеют линейную взаимосвязь.

2. Для сравнения эффективности проведения аукционов в зависимости от дня недели публикации объявления об аукционе, рассчитаны соответствующие данные, приведённые в таблице 1 (ещё на 3 аукциона заявки размещены в субботу).

Таблица 1

Эффективность проведения аукциона в зависимости от дня публикации

День публикации	Число запланированных аукционов	Процент состоявшихся аукционов
Понедельник	989	55,21 %
Вторник	921	60,59 %
Среда	701	56,49 %
Четверг	497	60,36 %
Пятница	779	59,82 %

Составлено авторами

Убедившись с помощью критерия Барлетта в однородности дисперсий, были проведены попарные проверки гипотез о равенстве средних значений. В результате проверок установлено, что различие в процентах состоявшихся аукционов между двумя группами (понедельник, среда и вторник, четверг, пятница) значимо (с уровнем значимости 0,05). Таким образом, можно предположить, что день публикации объявления об аукционе влияет на эффективность аукционов.

Аналогично проведено сравнение эффективности проведения аукционов в зависимости от дня недели окончания приема заявок. Данные приведены в таблице.

Таблица 2

Эффективность проведения аукциона в зависимости от дня окончания приёма заявок

День окончания приема заявок	Число запланированных аукционов	Процент состоявшихся аукционов
Понедельник	989	59,76 %
Вторник	921	48,39 %
Среда	701	57,70 %
Четверг	497	59,24 %
Пятница	779	60,00 %

Составлено авторами

Убедившись с помощью критерия Барлетта в однородности дисперсий, были проведены попарные проверки гипотез о равенстве средних значений. Статистически значимо оказалось отличие эффективности во вторник от эффективности в другие дни недели.

3. Для выявления важности признаков, направления связи между признаками и результатом аукциона были построены две модели логистической регрессии. В первой модели осуществлялось прогнозирование результата аукциона от всех переменных, во второй – из числа переменных был исключены коды ОКПД (из-за их многочисленности). Для факторных переменных «дни недели» за базовый уровень были приняты значения в понедельник.

Проверка на тестовом множестве показала, что модель без учета кодов является более точной, поэтому для интерпретации использовалась именно она.

Значения коэффициентов β_i модели логистической регрессии и соответствующая вероятность ошибки приведены в таблице 3.

Таблица 3

Коэффициенты и значимость логистической регрессии

Переменная	Значение коэффициентов	Вероятность ошибки первого рода
β_0	-0,3874	0,05993
Оформление	0,03922	0,00426
Размещение Вт	0,02734	0,78686
Размещение Ср	-0,1033	0,33941
Размещение Чт	0,09389	0,52504
Размещение Пт	0,09313	0,49611
Размещение Сб	-13,16	0,95923
Окончание Вт	-0,3431	0,02103
Окончание Ср	-0,02311	0,89015
Окончание Чт	0,1598	0,20025
Окончание Пт	0,2431	0,21398
Начальная цена	$-5,439 \cdot 10^{-8}$	0,00164
Совместный	0,4461	0,00166
Для СМП	0,7536	$< 2 \cdot 10^{-16}$
Для УИС	2,363	0,00142
Инвалиды	0,6972	0,00264
Преференции	0,2991	0,00002

Составлено авторами

Если выбрать уровень значимости 0,05 (например, такой уровень по умолчанию принят в Microsoft Excel), то можно сделать вывод, что на результат не влияет день размещения заявок (вероятность ошибки превосходит уровень значимости), а для дней окончания срока подачи заявок важно только, чтобы он не приходился на вторник. Остальные параметры (число дней на оформление заявки, является аукцион совместным, является ли он аукционом для субъектов малого предпринимательства, предоставлялись ли преимущества учреждениям и предприятиям уголовно-исполнительной системы, организациям инвалидов, предоставлялись ли преференции) являются весьма важными (соответствующие вероятности значительно меньше уровня значимости). Значения коэффициентов можно трактовать как изменение логарифма отношения шансов при увеличении на единицу значения входной переменной [9]. Знак перед коэффициентом показывает направления изменения.

4. Регрессионный анализ с помощью модели CART позволил разделить коды покупаемых товаров и услуг на четыре множеств с различной средней эффективностью аукционов. Примеры предметов аукционов и средняя эффективность приведены в таблице 4.

Таблица 4

Множества с различной средней эффективностью аукционов

№	Число аукционов в множестве	Пример предмета аукциона	Средняя эффективность
1	532	Оказание услуг по вывозу твердых коммунальных отходов	16 %
2	390	Приобретение игрового оборудования и инвентаря	38 %
3	2063	Поставка лекарственных препаратов	60 %
4	905	Поставка линолеума	87 %

Составлено авторами

Использование модели CART как классификатора не дало удовлетворительных результатов, поскольку определяющими признаками являются именно коды ОКПД. Вычисляемая системой R важность переменных для дерева решений [2] даёт значения 93 для кода ОКПД и максимум 2 для других переменных.

5. Использование случайных лесов для анализа эффективности было ограничено тем, что максимальное количество уровней факторной переменной ограничено в R числом 53

Не следует размещать заявки в понедельник (и, как следствие, заканчивать приём заявок во вторник). Удовлетворительного объяснения этим фактору на основе анализа дать нельзя.

Особое внимание следует обратить на те отрасли (коды ОКПД), где эффективность проведения аукционов мала. Возможно, торговым площадкам и заказчикам аукциона следует направлять информацию о заявках непосредственно предприятиям, которые могут быть потенциальными участниками.

Как следует из проведённого анализа, при привлечении большего числа участников и успешности проведения аукциона, можно рассчитывать на значительное снижение цены государственных закупок и, как итог, максимально эффективное распоряжение бюджетными средствами.

ЛИТЕРАТУРА

1. Breiman L. Random forests // Machine learning. – 2001. Т. 45. – № 1. – С. 5-32.
2. Ledolter J. Data mining and business analytics with R./J. Ledolter – John Wiley & Sons. – 2013. – 342 с.
3. Арепьев А.В. Типичные ошибки заказчиков при осуществлении государственных и муниципальных закупок по ФЗ-44 // Госзаказ: управление, размещение, обеспечение. – 2015. – № 41.
4. Гмурман, В.Е. Теория вероятностей и математическая статистика: учебник для прикладного бакалавриата / В.Е. Гмурман. – 12-е изд. – М.: Издательство Юрайт, 2018. – 479 с.
5. Введение в статистическое обучение с примерами на языке R./Г. Джеймс [и др.]. – М: ДМК Пресс, 2017. – 456 с.
6. Ильин А.А. Некоторые особенности размещения государственных и муниципальных заказов посредством проведения аукциона //Инновационная наука. – 2016. – № 2-3 (14).
7. Лепетикова И.Ю., Микаелян Д.С. Совершенствование системы государственного управления в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд // Ростовский научный журнал. – 2016. – № 11. – С. 75-83.
8. Маковлева Е.Е. Анализ практики государственных и муниципальных закупок у субъектов малого предпринимательства // Интернет-журнал Науковедение. – 2016. – Т. 8. – № 2 (33).
9. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям / Н.Б. Паклин, В.И. Орешков – СПб.: Питер, 2013. – 704 с.
10. Перерва О.Л., Степанов С.Е., Незимова С.С. Сравнение эконометрических моделей и методов бизнес-аналитики предсказания банкротства предприятий // Вестник евразийской науки. – 2017. – Т. 9. – № 6. – С. 74-74.

Pererva Olga Leonidovna

Bauman Moscow state technical university
Kaluga branch, Russia
E-mail: pol@bmstu-kaluga.ru

Stepanov Sergei Evgen'evich

Bauman Moscow state technical university
Kaluga branch, Russia
E-mail: stepanov@bmstu-kaluga.ru

Nezimova Sofia Sergeevna

Bauman Moscow state technical university
Kaluga branch, Russia
E-mail: stepanova_s@adm.kaluga.ru

Using big data analysis to determine the factors of the effectiveness of the public procurement process

Abstract. The article attempts to analyze the effectiveness of the auction from the point of view of the organizers. Analysis of auction statistics in electronic form shows that only about 60 % of them are held, so you often have to re-place applications. This leads to additional costs for organizers, customers, trading platforms and suppliers. The identification of factors affecting the effectiveness of auctions, and the development of recommendations to organizers, trading platform and customers is the purpose of this work.

Data on electronic auctions held in Kaluga region in 2017 in accordance with Federal law No. 44-FZ were taken for analysis. As a result of processing of initial data the investigated set consisted of 3890 records containing information on 13 variables. The authors used such methods of analysis as correlation analysis, statistical hypothesis testing, logistic regression, decision trees, random forests. The choice of these methods is explained both by the purpose of the research and their effectiveness.

The authors established the relationship between whether an auction was held or not and the factors under consideration. Using a combination of different models, the most important factors influencing the effectiveness of the auction are identified.

Based on the results of data analysis, certain conclusions were drawn and recommendations were made. For example, the range of the initial price at which the auction is most likely to take place is defined. The authors identified groups of goods and services with different efficiency of auctions, which can serve as a basis for work between customers, organizers and suppliers of goods and services.

Keywords: public procurement; electronic auctions; logistic regression; decision tree; random forests