

# Применение методов Data Science и математической статистики для структурирования спроса и предложения на промышленные товары и услуги

*Жукова Людмила Вячеславовна*  
старший преподаватель департамента прикладной экономики,  
Национальный исследовательский университет «Высшая школа экономики»,  
специалист ЗАО «ЕС-лизинг»,  
117587, город Москва, Варшавское шоссе, д. 125, стр.1, офис сек 4 эт 2 п VII к 5  
[lvzhukova@hse.ru](mailto:lvzhukova@hse.ru)

*Ковальчук Иван Михайлович*  
Старший бизнес-аналитик ЗАО "ЕС-лизинг"  
117587, город Москва, Варшавское шоссе, д. 125, стр.1, офис сек 4 эт 2 п VII к 5  
[ikovalchuk@ec-leasing.ru](mailto:ikovalchuk@ec-leasing.ru)

*Кiryushina Анастасия Алексеевна*  
К.э.н.,  
Старший бизнес-аналитик ЗАО "ЕС-лизинг"  
117587, город Москва, Варшавское шоссе, д. 125, стр.1, офис сек 4 эт 2 п VII к 5  
[akiryushina@ec-leasing.ru](mailto:akiryushina@ec-leasing.ru)

*Чугунов Виталий Рудольфович*  
ИТ директор ЗАО "ЕС-лизинг"  
117587, город Москва, Варшавское шоссе, д. 125, стр.1, офис сек 4 эт 2 п VII к 5  
[vhugunov@ec-leasing.ru](mailto:vhugunov@ec-leasing.ru)

**Аннотация:** в статье предлагается описание подхода к решению задачи структурирования спроса и предложения товаров и услуг. Предлагаемый подход, основанный на использовании методов Data Science, позволит реализовать современный инструментарий контроля за развитием промышленности в г. Москве. Подобный инструментарий дает возможность анализировать большое количество важнейших разнообразных структурированных, не структурированных и слабо структурированных данных из любых открытых источников, что позволит максимально точно оценить появление позитивных и негативных тенденций в изменении показателей деятельности предприятий промышленности г. Москвы.

**Ключевые слова:** кластеризация, Data Science, спрос и предложение, промышленные товары и услуги.

# Application of Data Science and mathematical statistics methods for structuring supply and demand for industrial goods and services

*Zhukova Lyudmila Vyacheslavovna*  
senior lecturer, Department of applied Economics,  
national research University Higher school of Economics»,  
specialist of JSC "EC-leasing»,  
117587, Moscow, Varshavskoye shosse 125, building 1, office 4 h at 2 p VII 5  
[lvzhukova@hse.ru](mailto:lvzhukova@hse.ru)

*Kovalchuk Ivan Mikhailovich*  
Senior business analyst, JSC "EC-leasing"  
117587, Moscow, Varshavskoye shosse 125, building 1, office 4 h at 2 p VII 5  
[ikovalchuk@ec-leasing.ru](mailto:ikovalchuk@ec-leasing.ru)

*Anastasia Kiryushina A.*  
Ph. D.,  
Senior business analyst, JSC "EC-leasing"  
117587, Moscow, Varshavskoye shosse 125, building 1, office 4 h at 2 p VII 5  
[akiryushina@ec-leasing.ru](mailto:akiryushina@ec-leasing.ru)

*Chugunov, Vitaly R.*  
Chief Information Officer (CIO) JSC "EC-leasing"  
117587, Moscow, Varshavskoye shosse 125, building 1, office 4 h at 2 p VII 5  
[vchugunov@ec-leasing.ru](mailto:vchugunov@ec-leasing.ru)

**Abstract:** The article describes an approach to solving the problem of structuring the supply and demand of goods and services. The proposed approach, based on the use of Data Science methods, will allow implementing modern tool for monitoring the development of industry in Moscow. Such tool helps to analyze a large number of structured, unstructured and poorly structured data from any open source. This allows to timely and accurately assess the appearance of positive and negative trends in the performance of industrial enterprises in Moscow.

**Keywords:** Data Science, clustering analysis, industrial goods and services, supply and demand

## Введение

В условиях цифровой трансформации, увеличения потоков информации и скорости экономических перемен крайне важно повышать качество контроля деятельности хозяйствующих субъектов, от производственно-экономических результатов которых зависит экономический потенциал города. Москва, как известно, один из самых крупных промышленно-экономических центров страны. Структурирование спроса и предложения на промышленные товары и услуги в этом городе создает основу для расстановки приоритетов мер поддержки развития предприятий промышленности города.

Авторами предлагается использование сквозной технологии больших данных для реализации инструментария дистанционного контроля за развитием промышленности в г. Москве с применением данных из внешних открытых источников, характеризующийся своим разнообразием и слабой структурированностью с которыми надо уметь

работать. Мониторинг и анализ большого количества данных, выполняемых с помощью методов Data Science, позволяет оценивать появление позитивных и негативных тенденций в изменении показателей деятельности предприятий промышленности г. Москвы.

Целью данного исследования является определить возможности структурирования спроса и предложения путем применения методов Data Science и математической статистики для решения задач в области контроля за развитием промышленности в г. Москве с применением данных из внешних открытых источников. Под этими данными понимаются интернет ресурсы, доступные каждому пользователю. Для исследования был выбран портал гос. закупок - сайт единой информационной системы в сфере закупок. Достижение цели исследования будет решаться на примере задачи структурирования спроса и предложения на промышленные товары и услуги.

Для достижения поставленной цели необходимо:

- сформулировать постановку задачи
- провести анализ предметной области;
- дать описание подхода к решению задачи;
- получить результаты, соответствующие требованиям.

## 1 Постановка задачи

Предлагаемый авторами подход решения задачи предполагает разработку и применение алгоритмов, позволяющих на основе анализа данных из внешних открытых источников решить следующие задачи:

- построить однородные группы предприятий г. Москвы по показателям, характеризующим спрос и предложение на промышленные товары и услуги по данным их деятельности на торговом портале государственных закупок;
- провести сравнительный анализ полученных групп на предмет сходства и отличий по финансово-экономическим показателям;
- сформировать перечень показателей, характеризующих спрос и предложение на группы промышленных товаров и услуг.

Подход решения задачи включает постановку цели, обследуемый период, объект исследования, данные, схему решения задачи и ожидаемые результаты (см. рисунок 1)

# Структурирование спроса и предложения на промышленные товары и услуги в г. Москве

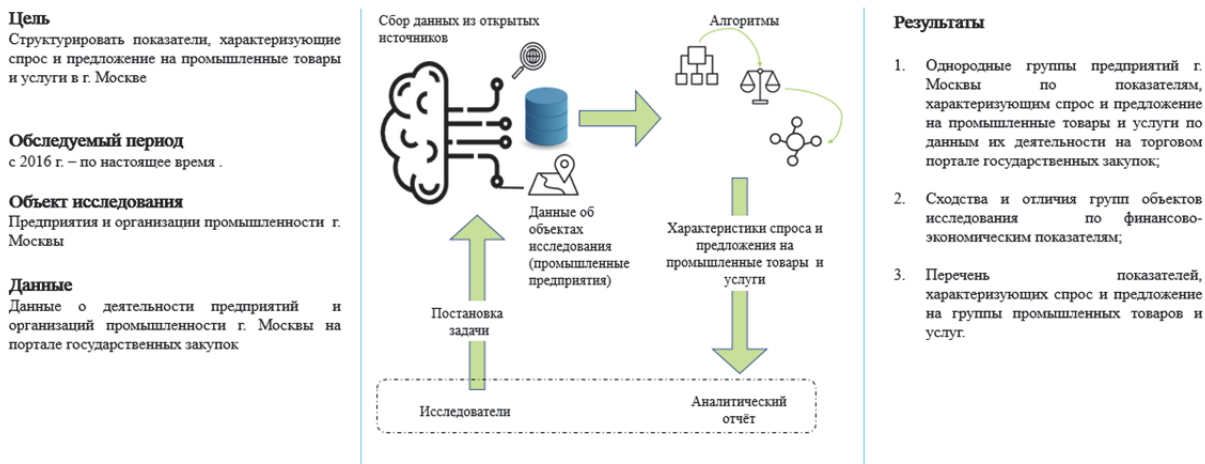


Рисунок 1 – Общая схема подхода решения задачи

По мнению авторов, использование предлагаемого подхода позволит повысить качество проведения мониторинга и анализа состояния экономики г. Москвы, а также даст объективную основу для совершенствования качества мероприятий по поддержке и стимулированию деятельности в сфере промышленности города Москвы.

Задача структурирования спроса и предложения на промышленные товары и услуги относится как к области промышленного маркетинга, так и к области государственного и муниципального управления развитием отрасли региона.

Исследование проведено с точки зрения необходимости расстановки приоритетов мер поддержки развития предприятий промышленности города. Одним из инструментов в аналитике при решении вопросов расстановки приоритетов мер поддержки развития предприятий промышленности города может стать структурирование спроса и предложения на промышленные товары и услуги.

Под структурированием в данном случае понимается набор однородных групп объектов исследования с характеристиками, отражающими особенности этих групп. Под спросом и предложением авторы понимают сделки между компаниями на интернет портале гос. закупок. Спросом являются тендеры, объявленные компаниями-заказчиками, а предложением – тендеры с установленными компаниями-исполнителями. Поэтому важно оценивать сделки между компаниями и непосредственно сами компании. В сегменте промышленных товаров и услуг на долю сделок с юридическими лицами приходится основной оборот, поэтому в статье изучаются рынки B2B, B2G, G2G, G2B.

Москва, как известно, является ведущим промышленным регионом России с высоким экономическим потенциалом. Объем ВРП на душу населения в 2016 году составил 1 157 тыс. рублей, что почти в 2 раза превышает среднероссийский уровень. На долю Москвы приходится более 1/3 объема валовой добавленной стоимости всех регионов России. Экономика Москвы характеризуется диверсифицированной постиндустриальной структурой ВРП. На долю промышленности приходится 16,1 % валовой добавленной стоимости (в том числе на обрабатывающие производства - 12%). [5]

Для получения показателей, характеризующих деятельность предприятий и организаций г. Москвы - производителей и поставщиков промышленных товаров и оказывающих услуги - авторами предлагается использовать по данным из открытых внешних источников данные с торговой площадки государственных закупок.

Чтобы проверить гипотезу о взаимосвязи характеристик торговой деятельности на площадке государственных закупок и характеристик спроса и предложения на промышленные товары проведен анализ расходов средств из государственного бюджета на предмет определения доли расходования бюджетных средств через торговую площадку государственных закупок. По результатам анализа выявлено следующее. Согласно данным из отчетов об исполнении федерального бюджета Российской Федерации и отчетов об исполнении бюджета г. Москвы на долю закупки товаров из бюджета приходится в среднем 21% всех бюджетных расходов, осуществляемых через торговую площадку государственных закупок. Это подтверждает значимость оборота через гос. закупки.

На сегодняшний день 90% закупок Москвы проходит в электронном виде. По данным “РТС-Тендер”, в 2017 г. российский рынок госзакупок достиг 36,5 трлн руб. На нем было заключено более 4 млн сделок. Лидером рынка является Москва – в 2017 г. столица закупила товаров и услуг на 2 трлн руб. по 44-ФЗ и на 13,9 трлн руб. по 223-ФЗ, что в сумме составляет около 44% всего объема госзакупок страны.

На площадке представлены в качестве заказчиков и поставщиков по г. Москве более половины всех юридических лиц, зарегистрированных в г. Москве. Среди средних и крупных предприятий около 80% зарегистрированы на торговой площадке государственных закупок.

Это подтверждает представительность данных с торговой площадки для характеристики спроса и предложения.

Признаками структурирования спроса и предложения на промышленные товары и услуги г. Москвы являются показатели, характеризующие группы предприятий и организаций г. Москвы, поставляющих промышленные товары и оказывающие услуги в г. Москве. Для структурирования необходимо выявить профиль типовых игроков рынка, характеристики их финансовой деятельности.

В качестве объектов исследования выступают предприятия, участвующие на портале государственных закупок, предметом исследования являются их финансовые показатели, свидетельствующие о прибыльности и доходности деятельности.

## 2 Описание подхода к решению задачи

Структурирование предполагает разбиение на однородные группы объектов, характеризующиеся схожими показателями. В качестве показателей выбраны финансовые показатели деятельности компаний за 2016 - 2017 год, показатели участия компании на торгах на госзакупках в 2016 и 2017 году.

Для решения задачи разработана экономико-математическая модель (далее ЭММ) (см. рисунок 2)

- набор алгоритмов расчета показателей, необходимых для кластеризации и типизации деятельности предприятий, задействованных на портале государственных закупок.

Структура компаний промышленного сектора не однородна по размеру. Авторы отмечают особенности крупных компаний, отличающихся от остальных игроков рынка. В г. Москве действуют несколько крупных компаний промышленного сектора, являющихся выдающимися для рынка всей РФ, такие как АО Вимм-Билль-Дан, АО "МПО им. Румянцева".

Поэтому ЭММ, разработанная авторами, предполагает сначала типизацию объектов по размеру компаний, выделение крупных, “доминирующих” компаний, и сегментацию остальных на однородные группы.

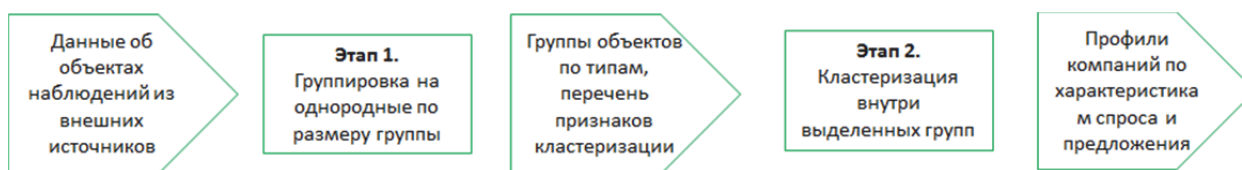


Рисунок 2 – Схема ЭММ

В общем подходе предусмотрено рассмотрение объектов и их типизация по заданным признакам, затем сегментация уже внутри выделенных групп. Тогда выделенные в процессе применения математических методов группы будут однородны по набору выявленных признаков.

Для типизации и сегментации были отобраны следующие существенные показатели деятельности компании, влияющие на возможные объемы его производства (как для характеристики предложения его продукции, так и для оценки спроса этого предприятия на продукцию и услуги других компаний промышленного сектора). Были выбраны две группы показателей:

1. Характеристики деятельности компаний в целом включают:
  - Показатели уставного капитала;
  - Средние показатели финансовой деятельности;
  - Анализ динамики выручки;
  - Капитал и валюта баланса;
2. Показатели участия в торгах на портале государственных закупок:
  - Доля участников в роли поставщика, %
  - Среднее количество контрактов
  - Средняя сумма контракта, млн. рублей

Статистика и эконометрика предполагает только построение экономико-математических моделей в то время, как Методы Data Science позволяют автоматизировать проверку гипотез, предобрабатывать данные, выявляя аномалии, а также изучать структуру данных.

Кластерный анализ как метод получения групп является разделом машинного обучения, который является одним из разделов Data Science

По результатам кластеризации определяются диапазоны значений в каждом типовом кластере, характеризующем предприятия, создающие спрос предприятия, создающие предложение, которые могут быть использованы в разработке шаблонов управленческих решений в области управления мерами поддержки и стимулирования деятельности в сфере промышленности предприятий среднего, малого бизнеса и микропредприятий.

В решаемой задаче применялась именно кластеризация, так как сущность полученных классов заранее не известна.

Авторы статьи выбрали для апробации методы кластеризации - BIRCH (двухшаговая кластеризация) и самоорганизующиеся карты Кохонена.

При выборе метода кластеризации учитывалось, что показатели кластеризации по выборке все числовые, и есть необходимость выбрать число кластеров. Число объектов невелико (до 1000). Также выбранные методы, реализованные в стандартных прикладных программах статистической обработки, а также в бесплатных библиотеках машинного обучения, написанные для языка программирования Python.

Метод BIRCH (двухшаговая кластеризация) как раз основан на разбиении неоднородных объектов на группы по заданным числовым показателям. На первом шаге на основании критерия Акаике выбирается оптимальное число кластеров. Затем происходит последовательно разделение.

Данный метод относится к классу иерархических. Все алгоритмы кластеризации разделяются на два вида по способу обработки данных, а именно на иерархические и неиерархические. Особенность иерархических алгоритмов заключается в том, что они предполагают последовательный пересмотр всех расстояний и выбор наименьшего, после чего происходит объединение в кластер и далее снова рассматривается расстояние от созданного кластера до каждого объекта.

В качестве математического критерия качества построенной кластеризации методом BIRCH выбран сводный показатель силуэтной меры (более 0,6, максимум достигается – в 1). При таком уровне сохраняется баланс между количеством кластеров, значимостью различия между ними и точностью результатов.

Этот показатель показывает, насколько среднее расстояние до объектов своего кластера отличается от среднего расстояния до объектов других кластеров. Данная величина лежит в диапазоне [-1; 1]. Значения, близкие к -1, соответствуют плохим (разрозненным) кластеризациям, значения, близкие к нулю, говорят о том, что кластеры пересекаются и накладываются друг на друга, значения, близкие к 1, соответствуют "плотным" четко выделенным кластерам.

Достоинством метода является автоматизированная процедура выбора количества кластеров, высокая скорость и хорошая масштабируемость для любого количества точек. Недостатком этого метода является работа только с числовыми показателями, требует устранения пропусков в данных. Этот алгоритм при реализации содержит ошибки, обусловленные тем, что для добавляемой точки находится не самый близкий узел [4].

Альтернативный метод, основанный на другом подходе - карты Кохонена - на создании нейронной сети, весовых коэффициентах и наглядной визуализации полученных результатов.

Самоорганизующиеся карты (Self Organizing Maps, или SOM) – это один из видов нейросетевых [2, 8] алгоритмов, основная особенность которого состоит в том, что он подразумевает обучение без учителя, то есть не требуется априорной информации о принадлежности объектов к группе. Самоорганизующиеся карты часто используются в моделировании, прогнозировании, нахождении взаимосвязей и закономерностей в условиях большого объема данных. SOM представляет собой двухслойную сеть. Каждый нейрон первого (распределительного) слоя соединен со всеми нейронами второго (выходного) слоя, которые расположены в виде двумерной решетки.

Нейроны выходного слоя называются кластерными элементами, их количество определяет максимальное количество групп, на которые система может разделить входные данные.

Достоинства: используется универсальный аппроксиматор – нейронная сеть, обучение сети без учителя, самоорганизация сети, простота реализации, гарантированное получение ответа после прохождения данных по слоям. Недостатки: работа только с числовыми данными, минимизация размеров сети, необходимо задавать количество кластеров.

Таким образом, применение методов Data Science позволит получить набор однородных по своим показателям группы предприятий, характеризующие особенности спроса и предложения, типовые профили компаний, создающих спрос, компаний, создающих предложение и компаний, создающих и спрос, и предложение.

## 2.1 Описание анализируемых данных

Для исследования был сформирован список из 506 компаний, по которым доступны сведения в открытых источниках. Для анализа было выбрано 338 компаний г. Москвы, действующих в области промышленности как непосредственной в обрабатывающем производстве, так и косвенно относящиеся к промышленности, например, когда речь идет об оказании услуг, связанных с научной, инженерно-технической и профессиональной деятельностью.

## 2.2 Описание решения задачи

Задача решена в два этапа. На этапе 1 были отобраны 34 компании, это группа доминирующих компаний, которая получена путем отбора объектов исследования на основании прямых и косвенных критериев.

Прямыми критериями отбора являются показатели выручки, прибыли (убытка) до налогообложения и уставного капитала по данным бухгалтерской (финансовой) отчетности за 2016 год и за 2017 год.

Косвенными критериями отбора являются показатели влияния на отрасль согласно Приказу Минпромторга России от 13.04.2015 N 798 (ред. от 30.07.2018) "Об утверждении Перечня организаций, оказывающих существенное влияние на отрасли промышленности и торговли" и показатель максимальной суммы контракта согласно Федеральному закону от 05.04.2013 N 44-ФЗ (ред. от 01.05.2019) "О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд" (с изм. и доп., вступ. в силу с 12.05.2019).

Методика отбора в 1-ю группу заключается:

- в проверке превышения хотя бы одного из прямых критериев порогового значения:
  - o выручки более 600 млн. рублей в год;
  - o прибыли (убыток) до налогообложения более 200 млн. рублей в год;
  - o уставного капитала более 700 млн. рублей в год;
- с учетом косвенных критериев:
  - o влияния на отрасль;
  - o максимальной суммы контракта.

Результаты формирования 1-й группы используются в решении задачи группировки всех объектов и выделение групп по размеру.

По итогам сравнения результатов деятельности доминирующих компаний с показателями индекса промышленного производства г. Москвы наблюдается соответствие положительного тренда в промышленности и в обрабатывающем производстве по всему городу положительному тренду в финансовых результатах доминирующих компаний. Динамика налоговых платежей в 2016 - 2017 гг. представлена в таблице 1.

Таблица 1 – Динамика налоговых платежей в 2016 - 2017 гг.

Показатели	Темп роста 2016-2017 гг.	Темп прироста 2016-2017 гг.
Динамика налоговых доходов в бюджет г. Москвы	115%	15%
Динамика налоговых платежей доминирующих компаний	244%	144%

Показатели участия доминирующих компаний в контрактах на портале государственных закупок с исполнением в г. Москве представлена в таблице 2.

Таблица 2 – Показатели участия доминирующих компаний в контрактах на портале государственных закупок с исполнением в г. Москве

Показатели	2016 г.	2017 г.
Общее количество контрактов доминирующих компаний на портале государственных закупок, исполненных на территории г. Москвы	3 202	3 725
Общее количество контрактов на портале государственных закупок, на территории г. Москвы	195 000	220 000

Эти предприятия и организации составляют отдельную группу, не сопоставимую с остальными. Объем закупок этих компаний составляет почти 2% от всего количества контрактов по г. Москве с бюджетными и внебюджетными средствами.

Эти компании выступают как заказчики, так и поставщики. учитывая характеристики их финансовых показателей и количество контрактов, можно отметить их существенное влияние на спрос так и на предложение товаров и услуг промышленного сектора.

На этапе 2 рассмотрена структура оставшихся 304 организаций.

Количество объектов исследования в разрезе основной экономической деятельности представлены в таблице 3.

Таблица 3 – Количество объектов исследования по основной экономической деятельности

Наименование вида экономической деятельности объектов наблюдения	Количество объектов наблюдения	Доля от общего количества
Продукция обрабатывающих производств	148	48,7
Услуги, связанные с научной, инженерно-технической и профессиональной деятельностью	97	31,9
Услуги в области информации и связи	32	10,5
Услуги, связанные с недвижимым имуществом	13	4,3
Услуги по оптовой и розничной торговле; услуги по ремонту автотранспортных средств и мотоциклов	6	2,0
Услуги административные и вспомогательные	2	0,7
Услуги финансовые и страховые	2	0,7
Водоснабжение; водоотведение, услуги по удалению и рекультивации отходов	1	0,3
Сооружения и строительные работы	1	0,3
Услуги транспорта и складского хозяйства	1	0,3
Электроэнергия, газ, пар и кондиционирование воздуха	1	0,3
Всего	304	100,0

Остальные 304 компании были проанализированы на наличие однородных подгрупп методами кластерного анализа.

### 3 Результаты применения различных методов кластеризации

#### 3.1 Результаты выявления однородных групп методами кластеризации BIRCH

Для выявления однородных групп по набору признаков, характеризующих размеры компаний, был применен метод кластерного анализа. В процессе были определены оптимальное количество кластеров и произведено разбиение по

следующим показателям, характеризующим торговлю на государственных закупках и основные финансовые показатели компании за 2016 - 2018 гг.:

- 1) Участие в роли поставщика и покупателя
- 2) Количество заключенных контрактов за 2016, 2017 и 2018 гг.
- 3) Сумма контрактов за 2016, 2017 и 2018 гг.
- 4) Сумма общего баланса по бюджетной отчетности компании за 2016 и 2017 гг.
- 5) Рентабельность деятельности в 2016, 2017 гг.
- 6) Средняя и максимальная сумма контракта, заключенная на электронном портале государственных закупок.

Для проведения кластеризации авторами использовалась бесплатная библиотека машинного обучения для языка программирования python - scikit-learn, библиотеки для визуализации, обработки и анализа данных - pandas и matplotlib.

Результаты применения двухшагового метода кластеризации (BIRCH) позволили определить оптимальное количество кластеров и группировку на 3 кластера (см. рисунок 3)

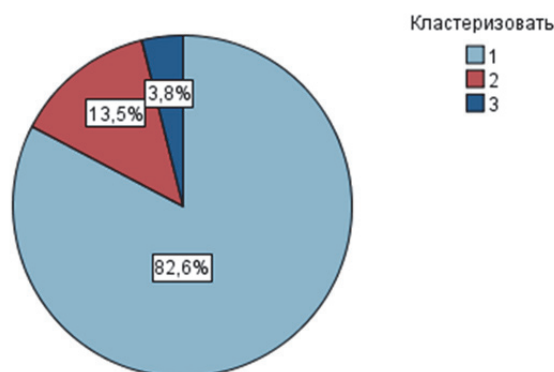


Рисунок 3 – Количественные характеристики кластеров, полученные двухшаговым методом кластеризации (BIRCH)

Основные характеристики кластеров представлены в таблице 4.

Таблица 4 – Основные характеристики кластеров

Наименование показателя	Кластер 1 (82,6%)	Кластер 2 (13,5%)	Кластер 3 (3,8%)
Уставной капитал, медианное значение (наиболее часто встречающееся), тыс. р	50	277 400	2 300 000
Средняя выручка, тыс.р.	900-1000	2800 - 3400	2 8000 - 30 000
Медианный (наиболее часто встречающийся) капитал, млн.р.	21 - 29	1700 - 2600	9100 - 9500
Средний прирост выручки, тыс.р.	90,54	613,19	2 490,57
Среднее количество контрактов на портале госзакупок	2 -5	71 - 80	300 - 440
Сумма контрактов на портале госзакупок, млн.р.	45 - 85	1141 - 1590	5500 - 6000

Распределение по отраслям представлено в таблице 5

Таблица 5 – Распределение по отраслям

Наименование вида экономической деятельности объектов наблюдения	Кластер 1 (82,6%)	Кластер 2 (13,5%)	Кластер 3 (3,8%)
Продукция обрабатывающих производств	51%	38%	55%
Услуги, связанные с научной, инженерно-технической и профессиональной деятельностью	29%	44%	45%
Услуги в области информации и связи	11%	5%	0%
Услуги, связанные с недвижимым имуществом	3%	13%	0%



Услуги по оптовой и розничной торговле; услуги по ремонту автотранспортных средств и мотоциклов	3%	0%	0%
Услуги административные и вспомогательные	1%	0%	0%
Услуги финансовые и страховые	1%	0%	0%
Водоснабжение; водоотведение, услуги по удалению и рекультивации отходов	0%	0%	0%
Сооружения и строительные работы	0%	0%	0%
Услуги транспорта и складского хозяйства	0%	0%	0%
Электроэнергия, газ, пар и кондиционирование воздуха	0%	0%	0%

Рассмотрим описание полученных кластеров:

1-й кластер

Наиболее многочисленный кластер (82,6% всех компаний) — это небольшие компании в среднем с уставным капиталом в 50 тыс.р., они составляют основную массу предложения товаров, выполняя небольшие по сумме контракты (6-9 млн), в среднем 2-5 штук в год. Эти компании участвуют на портале государственных закупок только в качестве поставщиков, удовлетворяя спрос на небольшие заказы. Т

2-й кластер

Кластер крупных компаний, выступающих чаще всего в качестве закупщиков, а не только поставщиков (87% компаний этого кластера), с уставным капиталом более 200 млн., имеющие выручку и капитал более 2 млрд, выполняющих около 80 контрактов в год на портале госзакупок.

Эти компании обеспечивают не только предложение, но и спрос на товары. Учитывая положительную динамику в финансовых показателях, можно отметить стабильный спрос с их стороны.

3-й кластер

Кластер из наиболее крупных компаний, на которые приходится большой оборот торговли, они выступают значимыми поставщиками на рынке промышленных товаров и услуг, имеют в среднем уставной капитал более 1 млрд р., положительные значения прироста выручки и капитала, и в среднем выполняют по 300 - 400 контрактов на портале госзакупок в год. Эти компании создают существенную часть спроса на продукцию промышленного сектора, выступая и со стороны предложения товаров и услуг.

### 3.2 Результаты выявления однородных групп с помощью карт Кохонена

На примере выборки из 450 предприятий (проект КСП промышленность) была решена задача кластеризации предприятий по группе финансовых показателей (баланс предприятия, прирост баланса, выручка, сумма контрактов на госзакупках и т. п.) и методом самоорганизующихся карт Кохонена (SOM). Этот метод на основе заданного количества итераций (25 000 итераций) разбил всю выборку на 8 кластеров.

При анализе полученных кластеров было выявлено, что 5 из них являются малыми по объему значений (до 10 единиц). Два самых крупных кластера, как и в предыдущем методе, собрали в себя самые крупные компании и самые небольшие, являющиеся исполнителями по контрактам на портале госзакупок.

Недостатком полученного метода является необходимость удалять наблюдения с пропущенными значениями, например, при отсутствии сведений о балансе.

Полученная кластеризация имеет экономическое обоснование, однако не учитывает особенности влияния некоторых факторов через нелинейное воздействие. Поэтому был применен подход, основанный на кластеризации с помощью нейронных сетей - карт Кохонена и осуществлена визуализация результатов. В дальнейшем, если модель карт будет использоваться, то это поможет выявить аномалии. Аномалия выявляется тогда, когда в следующий раз в набор данных попадут объекты, не похожие ни на один из известных образцов.

В результате применения метода было получено достаточно большое количество кластеров для интерпретации, некоторые со схожими характеристиками. В целом было выявлено также группа крупных предприятий, выступающих в качестве поставщиков и заказчиков. Они создают как спрос, так и предложение, сохраняя стабильность. Также выделилось несколько кластеров из малых и микропредприятий, которые выступают поставщиками, и отличаются лишь динамикой финансовых показателей.

### 3.3 Сравнительный анализ методов по применению к поставленной задаче

Благодаря типизации и методам кластеризации были выявлены 3 основные группы компаний, характеризующие спрос и предложение:

- 1) спрос обеспечивают крупные и средние предприятия, с сохранением финансовой стабильности спроса в 2016 - 2018 гг. Их доля относительно остальных не велика (25%), но размер выручки, прибыли и валюты баланса во много раз превосходит средние значения малых предприятий.

- 2) предложение создают малые и микропредприятия., динамика которых не стабильна, отмечается небольшая отрицательная динамика финансовых показателей в этом секторе.

Применение разных методов кластеризации дали схожие результаты, однако применение метода birch оказалось проще и доступнее. У метода BIRCH существует хорошо описанная документация на официальном сайте бесплатно распространяемой библиотеки машинного обучения. Задача визуализация самоорганизующихся карт Кохонена оказалась нетривиальной. Использование самоорганизующихся карт потребует более глубокие знания программирования на языке python. Плюсом использования метода BIRCH является заранее не предопределенное количество кластеров, что уменьшает количество итераций применения этого метода. Лучший по показателю меры качества (критерию качества кластеризации) результат в самоорганизующихся картах дается при большем количестве кластеров, часть из которых с малым количеством объектов.

## 4 Заключение

Авторы в своей работе провели исследование на предмет определения возможностей применения методов Data Science и математической статистики для решения задач в области контроля за развитием промышленности в г. Москве с применением данных из внешних открытых источников на примере задачи структурирования спроса и предложения на промышленные товары и услуги. Были получены профили компаний, отражающие особенности спроса и предложения в группах (кластерах).

Для достижения поставленной цели сформулирована постановка задачи, проведен анализ предметной области, дано описание подхода к решению задачи и получены результаты согласно поставленной задаче.

По результатам анализа предметной области получено обоснование гипотезы значимости использования открытых данных о деятельности предприятий и организаций, размещенных на электронной торговой площадке государственных закупок.

Стоит заметить, что при применении данного подхода к исследованию спроса и предложения предприятий, расположенных в небольших городах, результаты могут быть нерелевантными, вследствие отсутствия достаточного количества наблюдений для проведения анализа. Для решения этой проблемы следует определить набор показателей, присущих городам, в которых расположены исследуемые предприятия.

Помимо проведения аналитики для «бизнеса», данный подход может использоваться как инструмент контроля за развитием промышленности в городах Российской Федерации. Этот инструмент поможет определять аномальные активности компаний, на которые следует обратить внимание, находить «слабые места» промышленных отраслей, чтобы, например, субсидировать компании работающих в этих направлениях. Это позволит создать полноценные инфраструктуры, обеспечить локализацию инновационных производств, а также создаст новые высококвалифицированные рабочие места.

Подход, предложенный авторами, позволил структурировать спрос и предложение по финансовым признакам и по характеристикам объектов исследования на электронной торговой площадке государственных закупок. Таким образом выделены особенности отрасли, ее конкурентные преимущества.

Для второго этапа ЭММ (см. рисунок 2) были выбраны следующие методы: иерархический метод (BIRCH) и самоорганизующиеся карты Кохонена и проведена структуризация и группировки статистических данных, учтены специфика промышленных предприятий и отраслей.

Результаты применения выбранных методов подтверждают большие различия в структуре компаний, создающих спрос и предложение. Так, выделена группа очень крупных компаний (Доминирующие компании, лидеры своего рынка), и кластер 3 - крупные компании, которые являются лидерами по закупкам и поставкам товаров, обеспечивающие заметную долю поставок. Это компании, работающие по направлениям “Производство обрабатывающих производств” и “Услуги, связанные с научной, инженерно-технической и профессиональной деятельностью”.

Создают спрос на товары, помимо государственных нужд, и компании из кластера 2 — это предприятия среднего размера, с положительной, но небольшой динамикой финансовых показателей. Они преимущественно относятся к направлениям “Производство обрабатывающих производств”, “Услуги, связанные с научной, инженерно-технической и профессиональной деятельностью” и “Услуги, связанные с недвижимым имуществом”

Основная масса спроса удовлетворяется предложением большого количества небольших разнородных компаний, выполняющих несколько контрактов в год по своему направлению. Однако стоит обратить внимание на формирование эффективной системы поддержки предприятий для стимулирования спроса. Это поможет небольшим компаниям расширять ассортимент услуг и товаров, также поможет увеличить объем выпускаемой продукции.

Полученные результаты могут быть рекомендованы к использованию при реализации современного инструментария контроля за развитием промышленности в г. Москве. Подобный инструментарий дает возможность анализировать большое количество важнейших разнообразных структурированных, не структурированных и слабо структурированных данных из любых открытых источников, что позволит максимально точно оценить появление позитивных и негативных тенденций в изменении показателей деятельности предприятий промышленности г. Москвы.

## Список использованной литературы

- [1] Гржибовский А. М. Выбор статистического критерия для проверки гипотез // Экология человека. 2008. №11. 48-57
- [2] Терехов В.А., Ефимов Д.В., Тюкин И.Ю. Нейросетевые системы управления: Учеб. Пособие для вузов - М.: Высш. шк. 2002. - 183 с.: ил.
- [3] J. Hertz, A. Krogh, and R.G. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley, Reading, Mass., 1991
- [4] Кутуков Д. С. Применение методов кластеризации для обработки новостного потока [Текст] // Технические науки: проблемы и перспективы: материалы Междунар. науч. конф. (г. Санкт-Петербург, март 2011 г.). — СПб.: Реноме, 2011. — С. 77-83. — URL <https://moluch.ru/conf/tech/archive/2/207/> (дата обращения: 09.10.2019).
- [5] Официальный сайт Территориального органа Федеральной службы государственной статистики по городу Москве. Раздел Валовой региональный продукт. [https://moscow.gks.ru/wps/wcm/connect/rosstat\\_ts/moscow/ru/statistics/grp/](https://moscow.gks.ru/wps/wcm/connect/rosstat_ts/moscow/ru/statistics/grp/) – дата обращения: 30.04.19
- [6] Официальный сайт министерства финансов РФ [https://www.minfin.ru/ru/performance/budget/federal\\_budget/budgeti/2018/](https://www.minfin.ru/ru/performance/budget/federal_budget/budgeti/2018/) – дата обращения: 30.04.19
- [7] Проект «Хабр» Обзор алгоритмов кластеризации числовых пространств данных <https://habr.com/ru/post/164417/> [Электронный ресурс] – дата обращения: 30.04.19
- [8] BaseGroup Labs ООО «Аналитические технологии» Сегментация данных как метод сравнительного анализа <https://basegroup.ru/community/articles/comparative-analysis> [Электронный ресурс] – дата обращения: 01.05.19