

УДК 338.2  
JEL M15, L15, M11

## **АНАЛИЗ ТЕХНОЛОГИЧЕСКИХ ТРЕНДОВ РАЗВИТИЯ КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ В УСЛОВИЯХ ЦИФРОВИЗАЦИИ ПРОИЗВОДСТВА**

**Пруднус А. А.**

Кафедра проектирования и технологии производства электронных устройств  
Московский государственный технический университет им. Баумана  
г. Москва

**Карпунин А. А.**

Кафедра проектирования и технологии производства электронных устройств  
Московский государственный технический университет им. Баумана  
г. Москва

**Власов А.И.**

Кафедра проектирования и технологии производства электронных устройств  
Московский государственный технический университет им. Баумана  
г. Москва

**Фатхутдинов Т.М.**

Кафедра проектирования и технологии производства электронных устройств  
Московский государственный технический университет им. Баумана  
г. Москва

**Узенков Д.А.**

Кафедра проектирования и технологии производства электронных устройств  
Московский государственный технический университет им. Баумана  
г. Москва

*Аннотация. Работа посвящена анализу технологических трендов, задающих направление развития современных промышленных предприятий. Основное внимание уделено применению технологий работы с большими данными и основным тенденциям современного рынка. Показаны области применения технологии. Проведен системный анализ проблем, с которыми сталкиваются современные промышленные системы. В рамках повсеместной диджитализации описаны основные области технологий Big Data, IoT, Machine Learning и их связи со смежными областями науки и промышленности. В заключении даны рекомендации по использованию трендовых технологий для построения единой информационной структуры организации.*

*Ключевые слова: цифровая трансформация промышленности, Big Data, технологические тренды, производственные системы, Индустрия 4.0, диджитализация, синхронные технологии обработки данных, IoT.*

## **ANALYSIS OF TECHNOLOGICAL TRENDS OF DEVELOPMENT OF CORPORATE INFORMATION SYSTEMS UNDER CONDITIONS OF DIGITALIZATION OF PRODUCTION**

**Prudius A.A.,**

Department of Design and Production Technology of Electronic Devices  
Moscow State Technical University Bauman  
Moscow, Russia

**Karpunin A.A.,**

Department of Design and Production Technology of Electronic Devices  
Moscow State Technical University Bauman  
Moscow, Russia

**Vlasov A.I.,**

Department of Design and Production Technology of Electronic Devices  
Moscow State Technical University Bauman  
Moscow, Russia

**Fathutdinov T.M.,**

Department of Design and Production Technology of Electronic Devices  
Moscow State Technical University Bauman  
Moscow, Russia

**Uzenkov D.A.,**

Department of Design and Production Technology of Electronic Devices  
Moscow State Technical University Bauman  
Moscow, Russia

*Annotation. The work is devoted to the analysis of technological trends that set the direction of development of modern industrial enterprises. The main attention is paid to the application of technologies for working with big data and the main trends of the modern market. Areas of application of technology are shown. A systematic analysis of the problems faced by modern industrial systems. In the framework of ubiquitous digitalization, the main areas of Big Data, IoT, Machine Learning technologies and their relationships with related fields of science and industry are described. In conclusion, recommendations are given on the use of trending technologies to build a unified information structure for an organization.*

*Keywords: digital transformation of industry, Big Data, technological trends, production systems, Industry 4.0, digitalization, synchronous data processing technologies, IoT.*

В современной динамичной и конкурентной производственной среде всё большее практическое применение находят элементы цифровизации корпоративных систем. Это предусматривает повсеместное внедрение киберфизических систем, предиктивную аналитику, применение технологий AR/VR (Augmented/Virtual Reality), внедрение элементов «интернета вещей» (IoT) и алгоритмов Машинного обучения (Machine Learning), использования децентрализованной архитектуры (применение технологии Blockchain) анализа и обработки Больших данных в сферу социально-экономических взаимодействий и производство. Данная концепция получила название "Индустрия 4.0" (основные положения сформулированы на Ганноверской выставке в 2011) [1].

В современной практике также возрастает роль управленческих, экономических, социально-психологических знаний [2], способствующих глубинному пониманию процессов трансформации современного промышленного предприятия.

В условиях современной экономике ключевую роль для корпоративных информационных систем играет умение анализировать и структурировать существующие потоки информации и в дальнейшем её использовать. Сегодня мир данных влияет на все аспекты жизненного цикла предприятий, включая людей. В настоящее время экстенсивный рост объёма информации приводит к появлению Больших данных как отдельного класса обрабатываемой информации, требующего использования сетевых вычислительных мощностей при обработке, поскольку человеку уже не под силу работа с такими объемами данных. Область применения Больших данных и в дальнейшем будет расширяться ввиду повсеместной диджитализации и, как следствие, накопления огромных объемов данных в научной, промышленной, транспортной и бизнес - сферах развития общества [3].

Как правило, каждая организация в процессе своей деятельности накапливает огромные массивы информации. Но доступ к этим информационным ресурсам часто затруднен, а поиск нужной информации занимает слишком долгое время. При проведении исследований следует учитывать, что комплексное решение проблем эффективного применения технологий Big Data направлено на создание, передачу и применение механизмов формализации и манипулирования информацией в единой семантической межпредметной иерархии, обеспечивающей единое и полное описание процессов динамики корпоративных информационных систем [4].

Целью работы является анализ современных методик использования Больших данных в финансовых и телекоммуникационных организациях для классификации действующих и потенциальных клиентских баз при помощи нахождения основных паттернов поведения клиентов.

Объектом исследований в данной работе являются особенности цифровой трансформации промышленности, на примере методов анализа информации о клиентах в современных банках и ТелеКом-компаниях.

*Обзор литературы.* Если говорить о возрастающей с каждым годом роли данных в построении нового типа организационных структур, необходимо помнить, что данные теперь являются неотъемлемой частью четвертой промышленной революции, в которой сейчас и существует современный мир. Данные, нейронные сети и многие другие трендовые технологии позволяют сегодня изменить традиционные процессы, снизить в них роль человека как обязательного элемента обработки информации.

Современный характер рассмотрения сложных социально-производственных систем определяет проблемы анализа протекающих процессов, трудоемкости построения информационных систем для формального описания, хранения и обработки знаний об их объектах и процессах. Поиск решения подобного рода проблем в настоящий момент построен вокруг использования синхронных технологий обработки данных в качестве основного инструмента генерации, хранения и обработки информации в любой из предметных областей.

Возвращаясь к роли данных на сегодняшний день, необходимо определиться, что же из себя эти данные представляют. Big Data в наиболее общем случае – это обозначение данных во всех возможных вариациях, эффективно обрабатываемых масштабируемыми программными инструментами. Координатное пространство, благодаря которому можно представить, что такое Big Data постоянно изменяется, создавая при этом отдельные группы терминов, практически не связанных друг с другом. В широком смысле принято говорить о «больших данных» как о социально-

экономическом феномене, связанном с появлением технологий для анализа и обработки огромных массивов данных, так что для наглядности зачастую этот термин представляют в виде облака тегов или mind map'a, объединяющего ключевые аспекты этого феномена (рис. 1).



Рисунок 1 – Ключевые понятия, объединенные термином Big Data

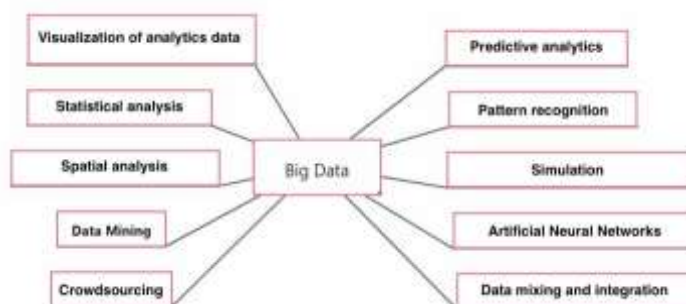


Рисунок 2 – Методы анализа Больших данных

При рассмотрении вопроса эффективного анализа данных и применения принципиально новых методов работы с ними необходимо использование наиболее передовых технологий для работы, ведь по всему миру используются методы статистического анализа, в котором с течением времени изменяется только набор инструментов [5]. Посредством использования продвинутой аналитики современные организации получают практическое применение общеизвестных статистических методов и как результат: глубинное понимание своей клиентской базы, особенности поведения отдельных групп людей, причинно-следственные связи их взаимодействия с бизнесом и политику, действующую на рынке в целом.

Анализ Больших данных с точки зрения основных характеристик можно определить так называемым «правилом VVV» (volume, velocity, variety) - объем, скорость и многообразие данных, рассматриваемых в том или ином случае. Классическими источниками таких данных можно считать интернет вещей (IoT) и социальные медиа [6], а также внутреннюю информацию организаций, получаемую в ходе работы на рынке (методы анализа представлены на рисунке 2).

Повсеместное внедрение цифрового инструментального производства в различные сферы деятельности и досуга активизирует деятельность малого бизнеса и индивидуального предпринимательства [7], при этом ИП сталкиваются с проблемой эффективности распределения ресурсов, актуальной является разработка индивидуального подхода в формировании условий работы с каждым в индивидуальности, а не создание единого клише для упрощения бизнес-процессов. Несмотря на то, что ещё в 2015 году термин Big Data был вынесен за пределы «кривой Гартнера» (рис. 3), вокруг этого термина и сегодня сохраняется высочайшего уровня ажиотаж, заостряя внимание на Big Data-трансформации.

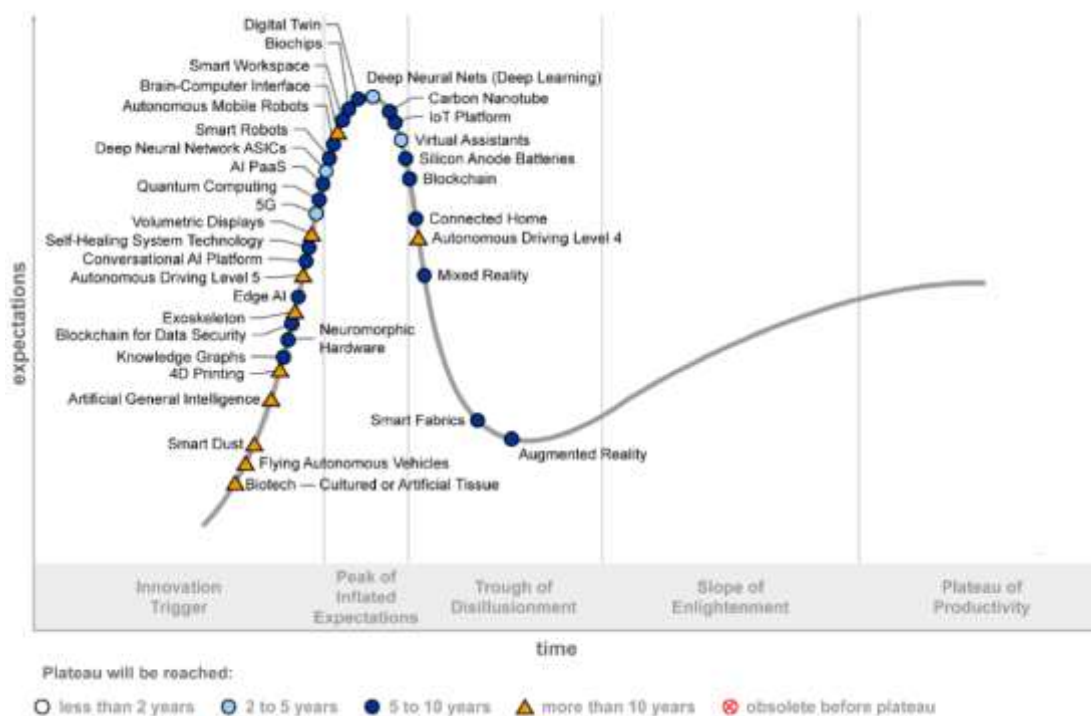


Рисунок 3 – «Кривая хайпа» исследовательской компании Gartner на 2018 год [8]

Но также в любом отдельном случае необходимо учитывать проблематику внедрения работы с Big Data, например, специфичность Big Data-проектов и целесообразность работы с такими объемами информации.

*Теоретическая часть.* Одним из ключевых понятий, связанным с Big Data, можно считать стратегию данных [9]. На сегодняшний день существует огромное количество мнений, что же из себя представляет стратегия данных, причем по-разному это понятие может быть воспринято как внутри одной организации, так и среди ключевых игроков рынка, производителей программного обеспечения и архитекторов данных. Но во всех взглядах есть общее ядро, которое разделяется каждым из участников и является одним из ключевых факторов выбора и реализации стратегии – это понимание цикла работы с данными. Здесь и появляется понятие жизненного цикла данных - основных этапов, проходимых данными в ходе ежедневного использования.

Одной из популярных моделей ЖЦ данных является модель Малькольма Чисхолма [10], в которой выделяется семь активных фаз взаимодействия с данными:

1. Data Capture – создание или сбор значений данных, которые еще не существуют и никогда не существовали в компании.

a. Data Acquisition – получение данных, предложенных внешними компаниями;

b. Data Entry – генерация данных вручную, при помощи ПО или физических устройств;

c. Signal Reception – получение данных с помощью Интернета вещей.

2. Data Maintenance – передача данных на серверы, где происходит синтез данных и их использование в форме, наиболее подходящей для этих целей.

3. Data Synthesis – создание ценности из данных с использованием логики, использование других данных в качестве входных данных.

4. Data Usage – применение данных как информации для задач, которые должны запускаться и выполняться внутри организации.

5. Data Publication – отправка данных клиентам или публикация в открытом доступе.

6. Data Archival – копирование данных для хранения на случай повторного использования и удаления из всех функционирующих производственных сред.

7. Data Purge – удаление каждой копии элемента данных внутри организации.

Если же рассматривать данные как основной актив, стратегия данных должна помогать использовать информацию внутри информационной системы, поэтому основными стейкхолдерами стратегии в первую очередь должны быть подразделения, увеличивающие активы организации [11]. Но необходимо помнить, что успех во взаимоотношениях с клиентами заключается, в том числе, в возможности уметь обработать информации о клиенте больше, чем клиент знает о себе сам (рис. 4).

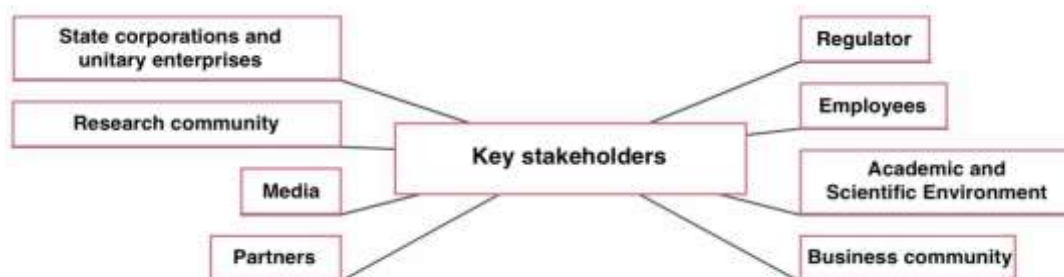


Рисунок 4 – Ключевые стейкхолдеры стратегии данных

У каждого этапа есть конкретный стейкхолдер, который может оказать наибольшее влияние на деятельность организации. Учет интересов большего числа стейкхолдеров позволяет снизить издержки на коммуникацию и хранение данных в будущем, а также повысить шанс на их монетизацию.

Говоря о базовых принципах обработки данных помимо правила “трех V” необходимо выделить горизонтальную масштабируемость систем работы с Big Data, подразумевающую под собой принцип разбиения системы на более мелкие структурные компоненты и разделение их на реальные физические машины с увеличением числа серверов, выполняющих в параллели один и тот же набор функций. Масштабируемость в таком контексте также подразумевает возможность внесения изменений в уже существующую физическую архитектуру, например, добавление новых серверов и увеличение перерабатывающих мощностей отдельных функциональных узлов для увеличения общей производительности системы.

Основными терминами, ассоциирующимися с понятием Big Data, являются NoSQL, Hadoop [12].

NoSQL - термин, принятый для определения семейства подходов, направленных на реализацию СУБД, имеющих существенные отличия от моделей, используемых в традиционных СУБД с доступом к данным с помощью языка SQL. Подход NoSQL применяется к базам данных, в которых делается попытка решить проблемы масштабируемости и доступности данных за счёт их атомарности и согласованности между собой.

Сама по себе идея ухода от реляционных баз данных появилась довольно давно, а использование нереляционных хранилищ началось ещё со времен появления первых ПК. В настоящий момент нереляционные базы данных нашли применение в специализированных хранилищах, например, иерархических службах каталогов. Появление нового поколения подобного рода систем произошло из-за необходимости создания параллельных распределённых систем для высокомасштабируемых интернет-приложений, таких как поисковые системы [13].

Hadoop является одной из основополагающих технологий работы с Большими данными. Вокруг этого проекта существует целая экосистема из связанных разработок и технологий, многие из которых развивались изначально в рамках проекта, а

впоследствии стали самостоятельными. Одной из основных целей Hadoop изначально было обеспечение горизонтальной масштабируемости кластера посредством добавления недорогих узлов (оборудования массового класса), без прибегания к мощным серверам и дорогим сетям хранения данных.

Одним из важнейших аспектов темы Big Data является качество данных. Качеством и проблемами в данных должны управлять специальные люди в организации, поскольку любые ошибки и неточности могут сильно исказить итоговые результаты работы с данными.

Первичным контролем должна выступать служба поддержки пользователей, которая приходит на помощь, если что-то случилось с их данными (сегодня такие специалисты называются дата-стюарды). Задача дата-стюардов - оперативно выявить причины проблем с данными так, чтобы ни один клиент не пострадал, или чтобы вовремя выпустилась отчетность.

За более детальную работу с проблемой качества данных отвечает более специфичная группа, называемая дата-инженеры. Специалисты такого рода проектируют единую архитектуру управления качеством данных (проверки, средства контроля и, конечно, дизайн самих IT-систем и цифровых интерфейсов).

Ключевыми характеристиками данных являются следующие параметры:

- Доступность – данные должны быть доступны для пользователя.
- Интерпретируемость – данные должны быть способны к интерпретации. К слову, не пытайтесь использовать мандаринский диалект, если вдруг пишете комментарии в проводках и так далее.
- Релевантность – данные должны быть релевантны для конечного пользователя, если они участвуют в процессе принятия решения.
- Точность – данные должны быть точны для пользователя, то есть быть точными и из достоверных источников.

*Обсуждение.* Большие данные все чаще находят пути развития в различных сферах общества, таких как телекоммуникационные технологии, банкинг, маркетинг, логистика. Примеры применения больших данных можно посмотреть в сфере телекоммуникационных технологий. В виду стагнации данной области рынка, новые технологии являются объектом пристального внимания и на изучение вариантов использования новых технологий выделяются большие финансы.

На примере компании МТС, победившей в премии CDO Award 2019 в номинации «За повышение эффективности бизнеса» (данная премия выдается за достижения в области работы с данными), можно рассмотреть наиболее удачные методы применения Больших данных на практике.

Одним из наиболее важным аспектом использования больших данных в сфере телекоммуникационных технологий является прогнозирование желаний клиента при помощи выявления предикторов в поведении клиента, благодаря данной технологии возможно предугадать желание клиента перейти на другого сотового оператора и предложить ему персональное предложение для его сохранения.

Так же еще одним не менее важным аспектом использования является управление персоналом - на сегодняшний момент были запущены проекты по управлению рабочим временем продавцов в салонах сотовой связи. Данная технология позволяет почасово прогнозировать на месяц вперед количество клиентов в каждом отдельном салоне связи по анализу транзакций данного отделения за последние 12 месяцев. Дополнительный экономический прирост приносит проект, позволяющий прогнозировать отток сотрудников, то есть выявляет вероятность увольнения по собственному желанию сотрудника в ближайший квартал. Данная технология

позволяет удержать высокооплачиваемых и дефицитных специалистов на нужных компании проектах.

В банковской сфере Большие данные применяются для оценки платежеспособности и надежности клиента при одобрении клиенту кредита. В данной области сырьем для обработки являются такие данные как цифровой след(социальные сети, форумы, история посещения различных сайтов и т.д.), информация из заявления на кредит и открытые источники информации( сайт судебных приставов, сайт налогообложения и т.д.). В результате обработки таких данных получается максимально отражающая действительность система оценки клиентов, в основе которой заложены статистические методы оценки.

С развитием data-аналитики сегодня всё большее количество компаний приходит к “облачным” решениям, применению Машинного обучения (ML) [14].

Одним из наиболее часто встречающихся практических кейсов в данном случае является использование алгоритмов машинного обучения для обработки Больших данных в сфере банкинга. Работа с клиентами зачастую требует знания большого количества факторов как о рынке в целом, так и о каждом из его участников по отдельности. Зачастую необходимо проводить системный анализ индивидуальных характеристик отдельных клиентов в рамках целой организации, чтобы улучшить те или иные показатели компании и, в дальнейшем, понимать, как завоевать лояльность потенциальных клиентов. При таком анализе выбирается как можно большее число показателей для построения наиболее эффективных моделей на предмет выбора управляющих воздействий поведением клиентов, причин возвращения клиентов и их распределение. В таких случаях аналитики, которую могут собрать стандартными методами, может быть недостаточно (рис. 5).

Полученная зависимость свидетельствует о том, что данные имеют различный уровень корреляции как между собой, так и с целевой переменной, относительно которой проводится анализ. Работа с такими данными требовала бы колоссальных временных затрат в случае отсутствия возможности машинной обработки.

Для подобного рода задач можно рассматривать такие модели работы с данными, как:

- Logistic Regression;
- Decision Trees;
- Random Forest;
- Deep Learning;
- Gradient Boosting;
- XGBoost.



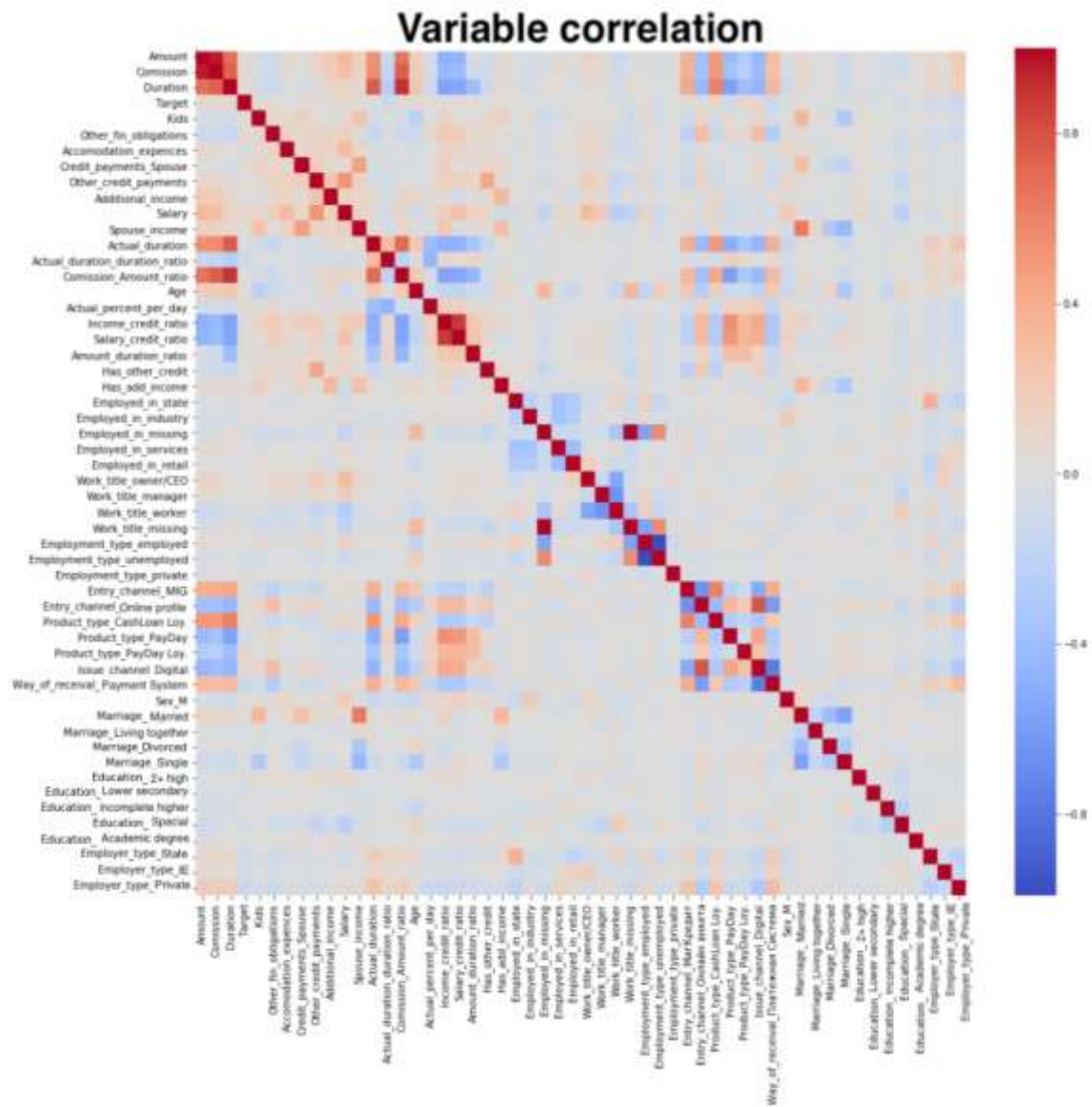


Рисунок 5 – Основные факторы, учитываемые при работе с клиентами в кредитных организациях

Наилучшим выбором является модель XGBoost, основанная на деревьях решений, использующая множественное количество деревьев (рис. 6.1, 6.2), обучающаяся на ошибках предсказаний в процессе итераций.

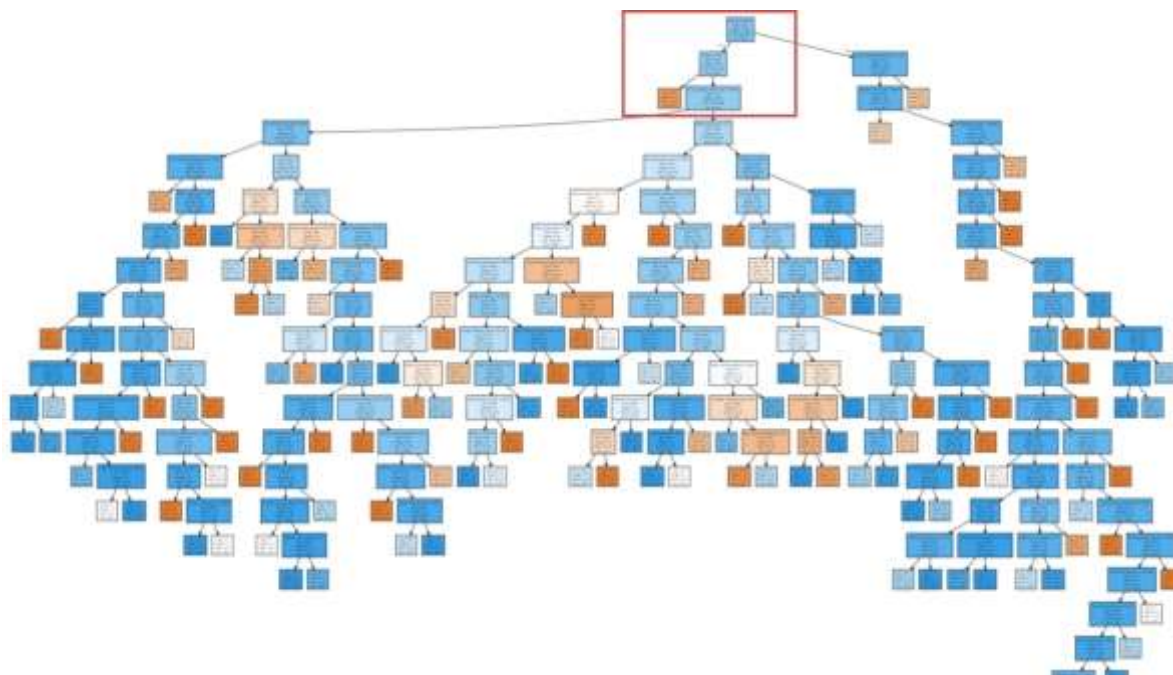


Рисунок 6.1 – Дерево решений выбранной модели работы с данными

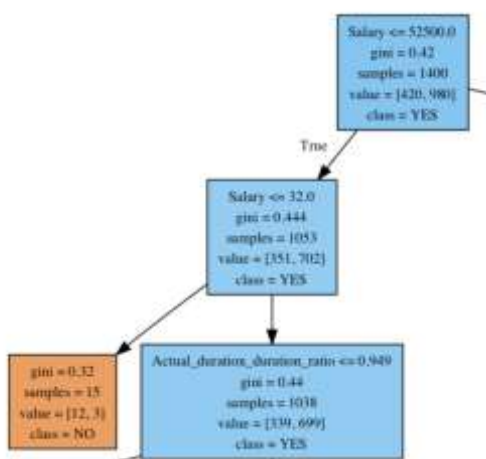
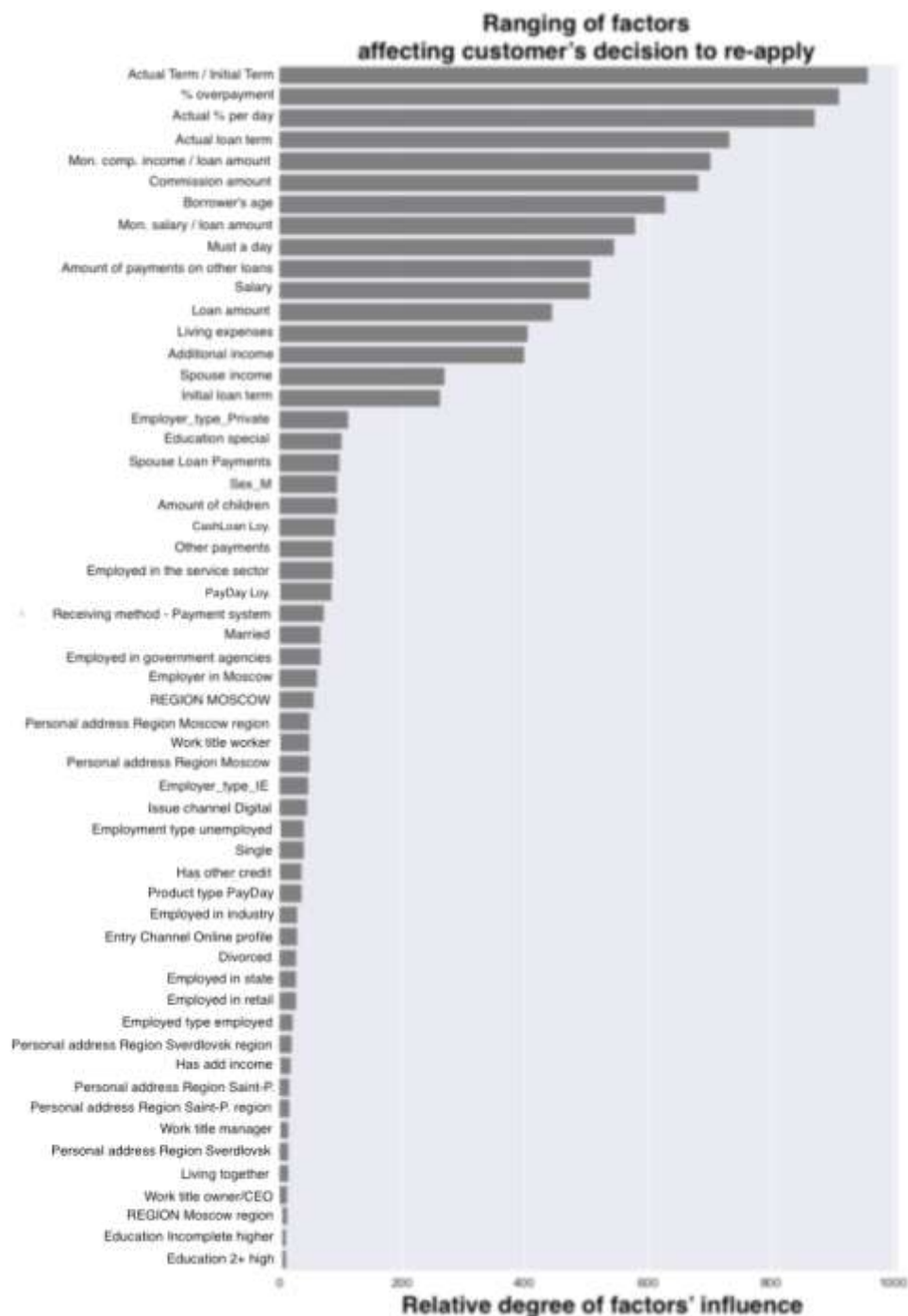


Рисунок 6.2 – Фрагмент дерева решений, выделенный на рис. 6.1

Ввиду большого количества переменных, линейного распределения не наблюдается. Укрупненные наборы параметров многократно разбиваются на подгруппы, удовлетворяющие определенным условиям, до окончательной классификации клиентов.

Ключевой вывод: «мануальное» использование результатов моделирования невозможно, для точной классификации и определения вероятностей необходимо использовать машинное обучение. Ввиду постоянного пополнения базы взаимодействия с клиентами, точность модели будет расти в каждой следующей итерации.

Результат анализа данных о клиентах при помощи алгоритмов машинного обучения может быть представлен в виде графических зависимостей (рис. 7), что является удобной формой представления для дальнейшей работы организаций.



*Рисунок 7 – Результаты анализа данных*

Конкуренция в современных условиях становится невозможной без применения углубленной аналитики и машинного обучения. Компании (в особенности финансовые организации), отставшие в данном направлении рискуют безвозвратно потерять долю рынка.

Машинное обучение остается одной из самых востребованных и внедряемых технологий [14]. И она еще не исчерпала свой потенциал. Не исключено, что ряд задач по интеграции источников данных также будет решаться с привлечением машинного обучения через анализ и интеграцию словарей (описание объектов данных в тех или иных источниках).

*Заключение.* Сегодня данные стали (или, по крайней мере, находятся в процессе становления) важной частью жизни человечества. Сервисы и продукты становятся цифровыми. Появляются новые инструменты и фреймворки [15], которые позволяют работать с данными максимально широкому кругу людей, так что крайне важно, чтобы люди понимали происходящие изменения и умели вести диалог о дальнейшем развитии в рамках цифровизации производства.

#### Список источников

1. Witten I. H. et al. Data Mining: Practical machine learning tools and techniques. – Morgan Kaufmann, 2016, С. 7-9.
2. Hastie, T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — 2nd ed. — Springer-Verlag, 2009. — 746 p. — ISBN 978-0-387-84857-0.
3. Berduygina O.N., Vlasov A.I., Kuzmin E.A. Investment capacity of the economy during the implementation of projects of public-private partnership // Investment Management and Financial Innovations. 2017. Т. 14. № 3. С. 189-198.
4. Vasconcelos D. R. et al. Mathematical model for a collaborative indoor position system (IPS) and movement detection of devices within IoT environment // Proceedings of the Symposium on Applied Computing. – ACM, 2017. – С. 602-608.
5. Vlasov A.I., Muraviev K.A., Prudius A.A., Uzenkov D.A. Load Balancing in Big Data Processing Systems // International Review of Automatic Control (I.R.E.A.CO.), 2019, Vol. 12, N. 1, pp. 42-47
6. Карпунин А.А., Власов А.И. Обработка данных с распределенным реестром в концепции "Индустрия 4.0" // В сборнике: Энергосбережение и эффективность в технических системах Материалы V Международной научно-технической конференции студентов, молодых учёных и специалистов. 2018. С. 120-121.
7. Williamson O. E. Transaction cost economics and organization theory // Industrial and Corporate Change. — 1993. — Vol. 2. — No. 2.
8. <https://www.gartner.com/en/newsroom/press-releases/2018-08-20-gartner-identifies-five-emerging-technology-trends-that-will-blur-the-lines-between-human-and-machine> [Электронный ресурс] – дата посещения 29.04.2019
9. <https://hbr.org/2017/05/whats-your-data-strategy> [Электронный ресурс] – Дата посещения – 29.04.2019
10. Благирев А. Big Data простым языком. – 2019
11. Akberdina V.V., Kalinina A.V., Vlasov A.I. Transformation stages of the russian industrial complex in the context of economy digitalization // Problems and Perspectives in Management. 2018. Т. 16. № 4. С. 201-211.
12. Shvachko K. Apache Hadoop. The Scalability Update — 2011. — Vol. 36, no. 3. — P. 7—13. — ISSN 1044-6397.
13. Demin A.A., Vlasov A.I., Visual methods of formalization of knowledge in the conditions of the synchronous technologies of system engineering // ACM International Conference Proceeding Series, 2017, N. 3166098, DOI: 10.1145/3166094.3166098.
14. Witten I.H., Frank E. Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). — Morgan Kaufmann, 2005 ISBN 0-12-088407-0
15. Yuldashev M.N., Vlasov A.I., Novikov A.N. Energy-efficient algorithm for classification of states of wireless sensor network using machine learning methods // Journal of Physics: Conference Series. 2018. Т. 1015. С. 032153.