



МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технологический университет «СТАНКИН»
(ФГБОУ ВО «МГТУ «СТАНКИН»)

ОТЧЕТ О ПРОХОЖДЕНИИ
ПРАКТИЧЕСКОЙ ПОДГОТОВКИ, ПРОВОДИМОЙ В ВИДЕ
Производственной практики (научно-исследовательская работа)

(вид практической подготовки)

ОБУЧАЮЩЕГОСЯ 3 КУРСА бакалавриата ГРУППЫ ИДБ-20-02
(уровень профессионального образования)

ЕРДОГАНА ДЕНИЗА ЕРДАЛОВИЧА

(ФИО полностью)

КАФЕДРА: Информационных технологий и вычислительных систем

НАПРАВЛЕНИЕ ПОДГОТОВКИ 09.03.01 Информатика и вычислительная техника
(СПЕЦИАЛЬНОСТЬ):

МЕСТО ПРОХОЖДЕНИЯ
ПРАКТИЧЕСКОЙ ПОДГОТОВКИ: ФГБОУ ВО «МГТУ «СТАНКИН»

СРОКИ ПРОХОЖДЕНИЯ 13.02.2023 – 11.06.2023
ПРАКТИЧЕСКОЙ ПОДГОТОВКИ:

МОСКВА
2023

Краткое содержание практической подготовки, проводимой в виде практики (проблемы и задачи, выбранные практикантом, способы их решения, полученные результаты, их оценка и самооценка, перечень компетенций формируемые в зависимости от программы обучения и вида практики)

Применение Data Science в финансовом секторе.

Посредством теоретического и практического исследования познакомиться с дисциплиной Data Science, дать её общую характеристику, рассмотреть её использование в различных прикладных экономических направлениях.

Изучение и анализ тематической литературы, поиск существующих исследований на тему использования.

На основании полученных данных был сделан вывод, что Data Science активно используется в финансовом сегменте.

Практикант Ердоган Дениз Ердалович
(ФИО полностью)

(подпись)

Отчёт о прохождении практической подготовки сдан « ____ » _____ 2023г.

Руководитель практической
подготовки от Университета Бекмурзаев В.А., к.т.н, доц.
(ФИО полностью, должность)

(подпись)



МИНОБРАЗОВАНИЯ РОССИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технологический университет «СТАНКИН»
(ФГБОУ ВО «МГТУ «СТАНКИН»)

ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ
для прохождения практической подготовки, проводимой в
ВИДЕ
Производственной практики (научно-исследовательская работа)

(вид практической подготовки)

ОБУЧАЮЩЕГОСЯ 3 КУРСА бакалавриата ГРУППЫ ИДБ-20-02
(уровень профессионального образования)

ЕРДОГАНА ДЕНИЗА ЕРДАЛОВИЧА

(ФИО полностью)

КАФЕДРА: Информационных технологий и вычислительных систем

НАПРАВЛЕНИЕ ПОДГОТОВКИ 09.03.01 Информатика и вычислительная техника
(СПЕЦИАЛЬНОСТЬ):

МЕСТО ПРОХОЖДЕНИЯ ПРАКТИЧЕСКОЙ ПОДГОТОВКИ: ФГБОУ ВО «МГТУ «СТАНКИН»

СРОКИ ПРОХОЖДЕНИЯ ПРАКТИЧЕСКОЙ ПОДГОТОВКИ: 13.02.2023 – 11.06.2023

МОСКВА
2023

1. Цель: Применение Data Science в финансовом секторе.
2. Задание: изучение основных терминов Data Science, применение их в экономической отрасли, анализ возможностей инструмента в будущем.
3. План-график выполнения работ:

№	Этапы	Сроки выполнения
1	Проанализировать литературу по темам «Data Science» и «Machine Learning».	13.02.2023 – 11.06.2023
2	Изучить основные термины Data Science	13.02.2023 – 11.06.2023
3	Дать определение терминам «Data Science», «Machine Learning», «данные», «информация» в рамках исследования.	13.02.2023 – 11.06.2023
4	Рассмотреть прикладное использование инструмента в финансовом сегменте.	13.02.2023 – 11.06.2023
5	Оценить перспективы развития направления связанного с данными.	13.02.2023 – 11.06.2023
6	Сделать вывод оценивающий полезность Data Science.	13.02.2023 – 11.06.2023

Практикант Ердоган Дениз Ердалович

(ФИО полностью)

(подпись)

Руководитель практической
подготовки от Университета

Бекмурзаев В.А., к.т.н, доц.

(ФИО полностью, должность)

(подпись)

Оглавление

ВВЕДЕНИЕ	6
ВСЁ ИЛИ НИЧЕГО О DATA SCIENCE	9
Математический аппарат	11
Дополнительные дисциплины необходимые в Data Science.....	16
ПЕРЕХОД DATA SCIENCE В РАЗНЫЕ ОБЛАСТИ	17
Data Science в экономике.....	19
DS в банковской структуре.....	19
DS в фондовых рынках и акциях.....	22
DS и социальные сети.....	23
ЗАКЛЮЧЕНИЕ.....	25
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ	26

ВВЕДЕНИЕ

Мы живём в информационной эре, главный ресурс на сегодняшний день – информация [1]. Все события происходят вокруг неё. Как говорил Натан Майер Ротшильд: “Кто обладает информацией, тот владеет миром” [2].

Информация – это данные имеющие ценность для субъекта. Данные — это совокупность сведений, зафиксированных на определённом носителе в форме, пригодной для постоянного хранения, передачи и обработки [17].

Ценность информации раскрывается при её использовании. Чтобы использовать информацию, её нужно: получить, и/или обработать, и/или сохранить, и/или передать [3].

Человеческий мозг может выполнить все основные операции при работе с информацией (перечисленные выше) [4]. Информационный объём мозга от 2,5 до 6 петабайта памяти [5], однако:

- Во-первых, большая часть памяти (нейронов в неокортексе) выделена под паттерны такие как: эмоциональные переживания, речь, двигательная активность, зрение, слух, сенсорные ощущения, которые нужны для базового функционирования человека [7].

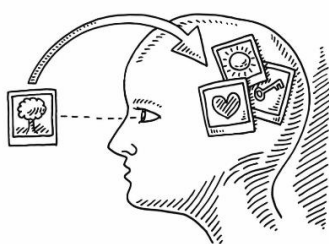
Результат, объём свободной памяти сильно снижается;

- Во-вторых, память человека разделена по временной структуре на:
 - мгновенную или сенсорную (работает до $\frac{1}{2}$ секунды);
 - кратковременную (сохраняет обобщенный образ в пределах 20 секунд);
 - оперативную (хранит данные на время решения поставленной задачи);
 - долговременную (позволяет вспоминать что-то спустя десятилетие) [7].

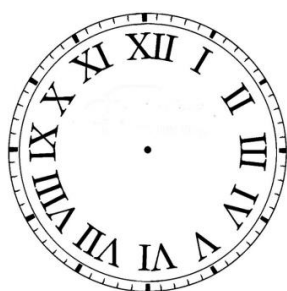
В связи с чем в трёх из четырёх случаях нарушается свойство полноты информации [8];

- В-третьих, человеческая память сама выбирает, что и когда запоминать. Что может отразиться на ценности информации [8];
- В-четвёртых, проблема информационного взрыва. Информационный взрыв — постоянное увеличение скорости и объёмов публикаций (объёма информации) [10]. Так в 2019 году объём трафика достигнул 2 зеттабайт в год.

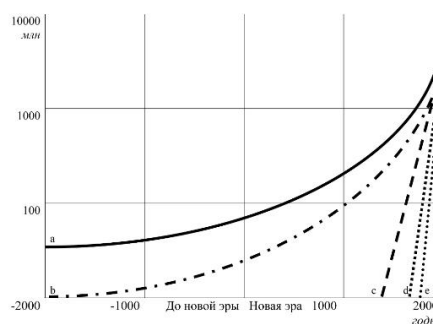
Темпы роста объёмов информации в разы больше возможностей мозга по обработке этой же информации. Это не все проблемы, которые существуют в данном вопросе.



**ПРОБЛЕМА ПРОИЗВОЛЬНОСТИ
ЗАПОМИНАНИЯ**



**ПРОБЛЕМА ВРЕМЕННОЙ
СТРУКТУРЫ**



**ПРОБЛЕМА ИНФОРМАЦИОННОГО
ВЗРЫВА**



ПРОБЛЕМА ПАТТЕРНОВ

Рис. 1. основные проблемы работы человека с информацией

На основе всего вышеперечисленного очевидно, что человеку не подвластны качественный контроль и использование информации без

посторонних инструментов. Тогда появляется спрос на инструменты для работы с информацией и данными. Данный вывод подчёркивает **актуальность темы**.

По закону спроса и предложения люди начинают искать необходимые инструменты [11]. Проблема хранения информации решается при помощи физических носителей информации и алгоритмов сжатия. Проблема обработки информации решается при помощи Big Data, Machine Learning и Artificial Intelligence (искусственный интеллект), которые объединяются в Data Science.

Big Data (большие данные) — это структурированные или неструктурированные массивы данных большого объема [18]. Machine Learning (машинное обучение) — это способ обучать компьютеры без программирования и явных инструкций, используя только шаблоны и логические выводы [19].

Объектом исследования является **Data Science** (наука о данных).

Предметом исследования является основы Data Science, его математический аппарат и применение вышеупомянутого в финансовом сегменте.

Цель исследования заключается в изучении принципов работы Data Science, его основных составляющих, а также выявление перспективы изучения данной науки.

Задачи исследования:

- 1) Проанализировать литературу по темам «Data Science» и «Machine Learning».
- 2) Изучить основные термины Data Science
- 3) Дать определение терминам «Data Science», «Machine Learning», «данные», «информация» в рамках исследования.
- 4) Рассмотреть прикладное использование инструмента в финансовом сегменте.
- 5) Оценить перспективы развития направления связанного с данными.

б) Сделать вывод оценивающий полезность Data Science.

Гипотеза, которая лежит в основе научно-исследовательской работы, состоит в предположении автора об большей эффективности использования инструментов Data Science в финансовом сегменте по сравнению с обычными аналитическими инструментами.

Этапы работы:

- 1) Поиск тематической литературы и других источников информации;
- 2) Изучение источников информации;
- 3) Анализ математического аппарата Data Science;
- 4) Написание работы.

В качестве **методов исследования** применялись: анализ теоретической литературы, синтез и изложение информации, изучение тематических медиафайлов.

Теоретическая ценность данного исследования обусловлена кратким и лаконичным изложением темы Data Science, их математического аппарата, прикладное применение инструмента в финансовом секторе.

Практическая ценность исследования заключается в полном изложении инструментов Data Science для оптимизации экономических задач.

В качестве **основных источников информации** использовалась русскоязычная и зарубежная техническая литература разных авторов, а также иностранные электронные журналы.

ВСЁ ИЛИ НИЧЕГО О DATA SCIENCE

Что же такое Data Science (наука о данных) — раздел информатики, изучающий проблемы анализа, обработки и представления данных в цифровой форме [12].

Data Science взаимодействует с большими данными. Большие данные — это огромные объёмы неструктурированной информации [13]. Ключевые слова здесь — «огромный объём» и «неструктурированность».

Основные задачи Data Science:

- Прогнозирование.

Прогнозирование — это процесс, при котором собранная информация используется для определения следующего действия пользователя. Точность в значительной степени зависит от количества доступных данных. Чем больше информации содержит модуль прогнозирования, тем лучшие результаты он может продемонстрировать. Пример: поле ввода текста в интернет-браузере. При вводе запроса в поисковике, появляются релевантные словосочетания для вас в раскрывающемся списке;

- Классификация.

Классификация - распознавание типа данных и группировка информации по этим типам. Пример: смартфон классифицирует изображения в одной папке, а видео в другой.

- Рекомендации.

Современные рекомендательные алгоритмы порой знают интересы и предпочтения пользователя лучше, чем он сам. Они выводят рекламу товаров, которыми вы когда-то интересовались и показывают вам ролики похожие на те, что вы недавно смотрели. Data Science изучает поведение человека, отслеживает закономерности и прогнозирует конечный результат, чтобы вывести наиболее оптимальную рекомендацию. Пример: рекомендация видеороликов в Youtube.

- Автоматизация.

Автоматизация позволяет снижать потребность в человеческом участии в работе, иницируя автоматизированный процесс какого-либо действия. Пример:

звонок будильника на вашем телефоне по утрам, тоже является автоматизацией, поскольку работает без непосредственного вмешательства человека.

- Машинное обучение.

Машинное обучение — это сбор данных, используемых частью технологии, наука о том, как обучить искусственный интеллект работать самостоятельно и расширять свои знания о мире.

- Искусственный интеллект.

Искусственный интеллект использует больше потоков данных для всестороннего сбора. Он позволяет технологиям учиться, прогнозировать и даже думать, используя для этого машинное обучение и другой пользовательский опыт.

- Аналитика данных.

Сегодня для достижения успеха в любом начинании стало незаменимым использование аналитики данных. Пример: служба поддержки напишет боту ответы на распространенные вопросы клиентов [15].

Все вышеперечисленные задачи основываются, в первую очередь, на математических методах.

Математический аппарат

Математика — это краеугольный камень Data Science. Хотя некоторые теоремы, аксиомы и формулы кажутся слишком абстрактными и далекими от практики, на самом деле без них невозможно по-настоящему глубоко анализировать и систематизировать огромные массивы данных. Ниже на рисунке 2 показаны основные разделы математики в часто используемые в дисциплине Data Science.

Математика для Data Science

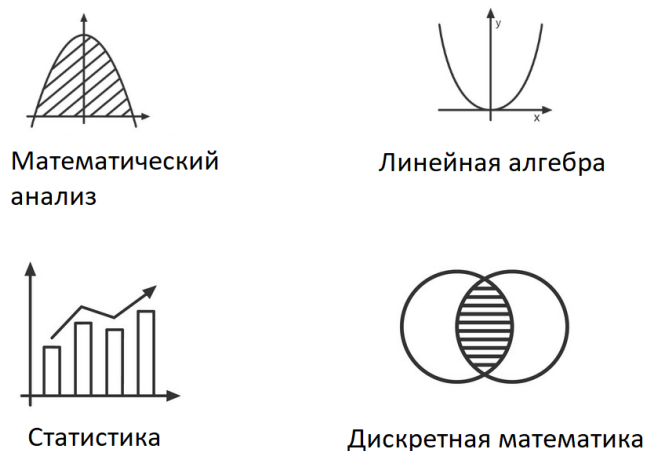


Рис. 2. Разделы математики используемые в Data Science

Базовой компетенцией специалиста по Data Science является его способность анализировать и интерпретировать данные, а математика — фундамент любой научной дисциплины. Почти все методы современной науки о данных, включая машинное обучение, имеют серьезную математическую основу.

Составляющие математической части:

- Классический математический анализ.

Дифференциальное и интегральное исчисление лежат в основе Data Science и машинного обучения. Они скрываются за простым на вид аналитическим решением обычной задачи наименьших квадратов в линейной регрессии или встроены в каждое обратное распространение, которое ваша нейронная сеть делает, чтобы узнать новый паттерн. Используемые темы:

- Функции одной переменной, пределы, лимиты, дифференцируемость;
- Теоремы о средних значениях, неопределенные формы, теорема Лопиталя;
- Максимумы и минимумы;

- Правило продукта и цепочки;
- Ряд Тейлора, концепции суммирования бесконечных рядов/интегрирования;
- Фундаментальные и средние значения, теоремы интегрального исчисления, оценка определенных и неправильных интегралов;
- Бета-и гамма-функции;
- Функции множественных переменных, предел, непрерывность, частные производные;
- Основы обыкновенных и дифференциальных уравнений в частных производных.

Применение: если вы когда-нибудь задумывались, как работает логистическая регрессия, то вам понадобятся знания о градиентном спуске и понятия из исчисления: производные, пределы и цепное правило.

- Линейная алгебра.

Это важный раздел математики для понимания того, как алгоритмы машинного обучения работают с потоком данных для создания представления. Все алгоритмы, лежащие в основе рекомендательных систем и распознавания изображений, включают в себя матрицы и матричную алгебру. Главные темы в данном разделе:

- Основные свойства матриц и векторов: скалярное умножение, линейное преобразование, транспонирование, сопряженность, ранг, детерминант;
- Внутренние и внешние произведения, правило умножения матриц и различные алгоритмы, обратные матрицы;
- Специальные матрицы: квадратная, единичная и треугольная матрицы, представление о разреженной и плотной матрице, единичные векторы, симметричная матрица, Эрмитова матрица, Антиэрмитова и унитарная матрицы;

- Концепция матричного факторизации, LU-разложение, Метод Гаусса, решение линейной системы уравнений $Ax = b$;
- Векторное пространство, базис, промежуток, ортогональность, ортонормированность, линейный наименьший квадрат;
- Собственный вектор, диагонализация, сингулярное разложение.

Применение: все нейросетевые алгоритмы применяют методы линейной алгебры для представления и обработки сетевых структур и операций обучения, а также для уменьшения размерности набора данных с наименьшим количеством параметров.

- Статистика.

Для Data Scientists важно понимание основных понятий статистики и теории вероятностей. Многие специалисты в этой области считают классическое машинное обучение ничем иным, как статистическим обучением. Важные концепции:

- Обобщение данных и описательная статистика, центральная тенденция, дисперсия, ковариация, корреляция;
- Базовая вероятность: основная идея, математическое ожидание, вероятностное исчисление, теорема Байеса, условная вероятность;
- Функции распределения вероятностей: равномерные, нормальные, биномиальные, хи-квадрат, t-распределение Стьюдента, центральная предельная теорема;
- Выборка, измерение, погрешность, генерация случайных чисел;
- Проверка гипотез, A/B testing, доверительные интервалы, p-values;
- ANOVA, t-тест;
- Линейная регрессия, регуляризация.

Применение: используются почти каждый день на этапе обработки и анализа данных.

- Дискретная математика.

Эта область математики не так часто обсуждается, но вся современная наука о данных использует вычислительные системы, и дискретная математика лежит в их основе. Повышение квалификации по дискретной математике будет включать в себя концепции, критически важные для применения алгоритмов и структур данных в аналитическом проекте:

- Наборы, подмножества, наборы мощности;
- Счетные функции, комбинаторика, счётность;
- Основные методы доказательства: индукция, доказательство противоречием;
- Основы индуктивной, дедуктивной и пропозициональной логики;
- Основные структуры данных: стеки, очереди, графики, массивы, хэш-таблицы, деревья;
- Свойства графа: связные компоненты, степень, концепции максимального потока/минимального разреза, раскраска графа;
- Рекуррентные соотношения и уравнения;
- Функция роста.

Применение: в анализе социальных сетей необходимо знать свойства графа, быстрый алгоритм поиска и обхода сети.

- Исследование операций.

Эти темы наиболее актуальны в специализированных областях, таких как теоретическая информатика, теория управления или исследование операций. Базовое понимание этих мощных методов также может быть полезным и в практике машинного обучения. Почти каждый алгоритм стремится минимизировать ошибку оценки с учетом различных ограничений, что является проблемой оптимизации. Основные темы:

- Основы оптимизации и как сформулировать задачу;
- Максимумы, минимумы, выпуклая функция, глобальное решение;
- Линейное программирование, Симплекс-метод; Линейное программирование, Симплекс-метод;
- Целочисленное программирование;
- Программирование в ограничениях, задача о рюкзаке (с англ. knapsack problem);
- Рандомизированные методы оптимизации.

Применение: простые задачи линейной регрессии с использованием функции наименьших квадратов потерь имеют точное аналитическое решение, но не задачи логистической регрессии. Для их решения необходимо понимать, как работает выпуклость в оптимизации.

Дополнительные дисциплины необходимые в Data Science

Без компьютерных программ и утилит обрабатывать сведения в электронном виде невозможно. Особенно тогда, когда речь идет о больших ее объемах. Многое зависит от того, какой именно специалист приступает к работе. Ориентироваться можно на следующие инструменты:

- 1) Все «ученые по обработке и аналитике материалов» должны разбираться в таблицах, СУБД, хранилищах, SQL и ETL.
- 2) BI-аналитик: инструментарий BI, SAS, R, Python, Knime, RapidMiner.
- 3) Специалистам по данным и «ученым»: библиотеки визуализации и проведения досконального анализа в R и Python, углубленное изучение майнинга, Docker, Airflow.
- 4) Инженерам: углубленные познания в ETL-процессах, а также в процессах выстраивания пайплайна.

Также необходимо иметь навыки в программирование. Обязательно знать SQL, а также Python. В идеале дополнить соответствующий багаж Scala и Java. В аналитике часто задействуются облачные платформы. Ещё рекомендуется использовать технологии обработки сведений в огромных объемах (Kafka, Hadoop, Spark).

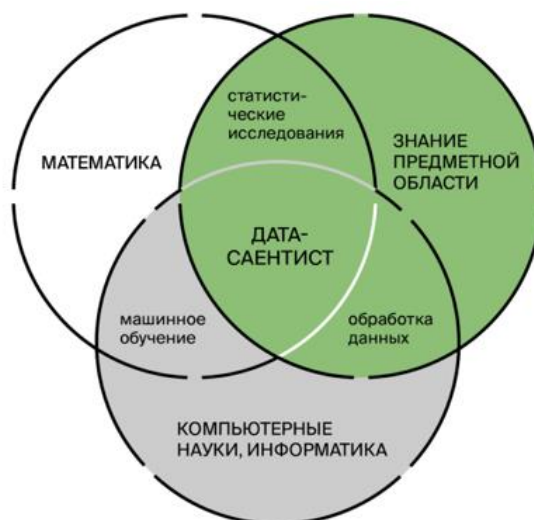


Рис. 3. Основные дисциплины в Data Science

ПЕРЕХОД DATA SCIENCE В РАЗНЫЕ ОБЛАСТИ

В наше время сильно востребованы IT-технологии. Предприниматели и владельцы компаний стремятся захватить часть рынка. Новые отрасли для бизнеса найти всё сложнее и сложнее, существует альтернативный подход конкурирования – улучшение старого.

Если рассмотреть все пиковые скачки популярнейших компаний на сегодняшний день, то можно заметить, внедрение новшеств прямо пропорционально росту акций. Так: так в 1977 году вышел apple 2 – первый персональный компьютер, Amazon в 2006 году запустил облачный сервис [19], [20].

2007	от 36.43 до 100.82 USD
2006	от 26.07 до 47.87 USD

Рис. 4. Изменение цены акций компании Amazon при выпуске облачных сервисов

Задача предпринимателей узнать как можно быстрее о трендах и новых инструментах с целью дальнейшего внедрения их в свои бизнесы для извлечения максимальной выгоды.

Относительно недавно (впервые термин появился в 2001 году, в дальнейшем развивался благодаря специалиста компании Facebook) data science стал популярным в медиа пространстве. Многие предприниматели сразу смекнули, что данный инструмент будет полезен в их делах. Логика проста, чем больше у нас информации, чем быстрее мы её обрабатываем, тем больше у нас возможностей.

В последнее время сфера Data science сильно развивается. На сегодняшний день DS проникает во все области: медицину, политику, экономику, банковские сети, айти компании, частные бизнесы. Но особый ажиотаж появился связанный с DS появился в финансовом секторе.



Рис. 5. Примеры использования Data Science (в частности Big Data)

Data Science в экономике

Компании обращаются к Data Science независимо от размера бизнеса. Рост спроса обусловлен тем, что предприятиям нужно анализировать огромные объемы информации, с которыми невозможно справиться без помощи ИИ и дата-сайнтистов. Согласно статистике от IDC и Hitachi, 78% респондентов столкнулись с увеличением потоков данных [22]. При этом в неструктурированной, на первый взгляд, информации могут быть ценные сведения, способные повлиять на результаты бизнеса.

В качестве примера, Apple Music, «Яндекс. Музыка», Spotify главной фишкой стали рекомендации. Если пользователь любит слушать рэп, ему никогда не предложат песни рок стиля. И подберут подходящих исполнителей, треки которых заставят человека остаться еще на 10–30 минут.

Подобные новшества приносят большие деньги бизнесам.

DS в банковской структуре

Банк представляет собой - компанию, которая занимается операциями с деньгами, ценными бумагами и драгоценными металлами. Она оказывает финансовые услуги государству, физическим и юридическим лицам.

Как и в любом деле связанным с услугами, исполнитель должен стараться удовлетворить запросы клиента, иначе клиенты не будут приносить деньги. Банки прекрасно это понимают. Поэтому банки начали использовать Data Science.

Способы использования DS банками:

- Сбор данных.

Они всё время собирают данные, то есть: как отреагировал клиент на рекламу, личное предложение, что клиент просматривает в личном приложении банка. Таким образом банк пытается дать максимально выгодное и

привлекательное предложение для клиента, а сам на этом заработать - бизнес ничего лишнего.

Очевидно, что чем быстрее и успешнее банковская система будет обнаруживать злоумышленников, тем лучше будет для клиентов, а как следствие будет выше рейтинг у банка.

- Обнаружение мошенничества.

Как уже было сказано, банковская система хорошо знает, как ведёт себя клиент, если в какой-то момент обнаруживается большое количество транзакций, новые виды операций, особенно, это актуально для опытных пользователей банков. В таком случае банковская система может попросить подтверждение действий при помощи указания кода с телефона/почты.

- Прогноз продолжительности жизни клиента.

Да, Вам не показалось, это действительно так, это объясняется тем, что величина жизненного цикла клиентов является прогнозом показателя прибыли, которую получит бизнес в результате взаимодействия с пользователями на протяжении жизни. Важность данной величины быстро растёт, так как она помогает создавать и поддерживать выгодные отношения с отдельными клиентами, что повышает прибыльность и рост бизнеса. Привлечение и удержание прибыльных клиентов является все более острой проблемой для банков. По мере усиления конкуренции, банки теперь нуждаются во всестороннем представлении каждого клиента, чтобы эффективно концентрировать свои ресурсы.

Поскольку необходимо учитывать, что требуется проанализировать большой объем данных, таких как: обзор новых и ушедших клиентов, использование разнообразных банковских продуктов и услуг, их объем и окупаемость, а также другие характеристики, включая географические, демографические и рыночные показатели. Кроме того, требуются постоянное

обновление, сортировка и прочие манипуляции с этими данными, чтобы они сохраняли свою значимость и актуальность.

- Сегментаций клиентов.

Data scient-исты выделяют всех клиентов в группы. У каждой группы есть свой коэффициент приоритета, над группами с большими коэффициентами проводится отдельная работа. В дальнейшем такая сегментация пользователей позволяет эффективно распределять маркетинговые ресурсы и подбирать индивидуальный подход к каждой группе клиентов. Стоит помнить, что сегментация клиентов призвана улучшить их обслуживание и помочь в формировании лояльности и удержании, что так необходимо для банковского сектора.

- Систему рекомендаций.

Инструменты data science и machine learning могут создавать простые алгоритмы, которые анализируют и фильтруют деятельность пользователя, чтобы сделать наиболее актуальное для него предложение. Такие механизмы с рекомендациями демонстрируют элементы, которые могут заинтересовать пользователя, даже до того, как он сам его искал. Чтобы создать механизм рекомендаций, data scientist'ы анализируют и обрабатывают много информации, идентифицируют профили клиентов и охватывают данные, показывающие их взаимодействия, чтобы избежать повторения предложений. Тип рекомендуемых систем зависит от метода фильтрации алгоритма. Совместные методы фильтрации могут быть основаны либо на пользователе, либо на элементах, и работать с поведением клиента для анализа предпочтений других пользователей, а затем давать рекомендации новому клиенту.

- Службы поддержки.

Выдающаяся служба поддержки - это ключ к сохранению продуктивных долгосрочных отношений с клиентами. В рамках обслуживания пользователей, поддержка является важной, но общей концепцией в банковской сфере.

Фактически, все банки являются сервисными предприятиями, поэтому большая часть их деятельности включает в себя элементы обслуживания, что представляет собой подробные и своевременные ответы на вопросы и жалобы клиентов, а также взаимодействие с ними. Data science делает этот процесс более автоматизированным, точным, продуктивным и менее затратным в отношении рабочего времени сотрудников.

DS в фондовых рынках и акциях

Наука о данных вошла в финансовые рынок и биржи. Например, современные хедж-фонды (инвестиционный фонд, ориентированный на максимизацию доходности при заданном риске или минимизацию рисков для заданной доходности) привлекают инвесторов умножением капитала при минимальном риске.

Хедж-фонды собирают деньги от всех инвесторов, на эти деньги покупают инструменты в определённом соотношении, в итоге при просадке одних активов, растут другие, в итоге мы получаем органическую систему, которая в временном хронометраже растёт.

Для достижения такой стратегии необходимо качественно анализировать тренды, и базисно прогнозировать котировки активов, что не так просто сделать. Раньше использовали стандартные инструменты аналитики: теханализ, графики, анализ общей тенденции моды.

Сейчас же с появлением новых инструментов, всё чаще встречается в распоряжении аналитиков DS-ные инструменты, которые позволяют менее затратно, точнее и эффективнее анализировать и прогнозировать данные фондов.

Среди приоритетных технологий для хедж-фондов — искусственный интеллект и машинное обучение. Проведенное ещё в 2018 году исследование BarclayHedge показало, что более половины игроков рынка используют эти методы для принятия инвестиционных решений и две трети — для генерации

торговых идей и оптимизации портфелей. Причем более половины использовали ИИ на протяжении трех лет, а треть — свыше пяти лет. Мало какая другая отрасль может похвастаться столь ранним использованием ИИ — пожалуй, кроме собственно IT-индустрии [22].

Такой интерес к искусственному интеллекту обусловлен необходимостью хедж-фондов быстро просчитывать тренды, искать в новостях сигналы о том, что в скором времени цена той или иной акции изменится. И чем точнее будут прогнозы, тем меньшие риски и выше прибыль.

Уже есть алгоритмы, которые позволяют агрегировать новости из различных деловых изданий и выявлять сигналы, способные изменить стоимость акции. Искусственный интеллект может анализировать гораздо большие объемы данных, чем человек, а скорость обработки информации у него гораздо выше. Чем больше данных проанализировано, тем точнее будет прогноз. Таким образом, ИИ выполняет задачи, которые поставил перед ним аналитик данные.

Data scientist изначально строит модель, основываясь на данных. Когда модель создана, она тестируется на предыдущих событиях — этот процесс называется *backtesting*. К примеру, есть модель, которая позволяет сделать прогноз на завтра при определенном поведении цен на акции. Этот прогноз ложится в основу действий хедж-фонда: именно он показывает, покупать или продавать конкретные акции. Чтобы оценить точность прогноза, в модель вводятся данные прошлых периодов. Например, можно ввести в систему данные, которые были собраны десять дней назад, сделать прогноз на день вперед и сравнить с тем, как вели себя цены на акции девять дней назад, заглянув таким образом в будущее.

DS и социальные сети

Социальные сети являются хорошим источником данных и важно уметь эффективно работать с этим данными. Стоит отметить, что существует

отдельное направление — Social Mining. Это применение методов и алгоритмов data mining для поиска и обнаружения зависимостей и знаний в социальных сетях. Области применения достаточно широки.

- Анализ инфопотоков.

Данный класс методов позволяет выявить лидеров мнений в соцсетях, осуществлять управление медиа-кампанией, оценивать отношение пользователей к той или иной информации. Решаемые задачи здесь такие:

- Поиск наиболее коммуницирующих между собой объектов;
- Поиск объектов, имеющих наибольшее количество связей;
- Поиск наиболее «авторитетных» объектов;
- Поиск объектов, которые служат «мостиком» между сообществами.

Одна из важнейших задач в анализе соцсетей — поиск «важных» участников социального графа. Для этого исследователи рассчитывают различные виды метрик:

- Degree centrality (по числу связанных узлов; важен тот, у кого много друзей; полезна для выделения лидеров мнений),
- Closeness centrality (по близости; на сколько близко участник ко всем другим в сети; чаще всего применяется в задаче поиска групп влияния и «серых кардиналов»),
- Betweenness centrality (по промежуточности; число кратчайших путей, проходящих через участника; насколько часто через этого человека проходит информация в сообществе).

- Анализ тональности.

Данный класс методов позволяет оценивать отношение пользователей к той или иной информации. Решаемые здесь задачи: оценка эмоциональной окраски сообщений; выделение именованных сущностей и оценка их эмоциональной окраски.

- Анализ тем.

Данный класс методов позволяет выявить темы, которые наиболее популярны в сообществе и чаще всего в нем обсуждаются. Решаемые задачи:

- выделение тем (topic modelling), оценка эмоциональной окраски тем, выделение сущностей, связанных с темой.

- Анализ изображений.

Позволяет выявить какие типы фото-контента размещают различные сегменты пользователей.

ЗАКЛЮЧЕНИЕ

Спрос на специалистов по Data Science в ближайшие пять лет будет только расти. Сайт Glassdoor четвертый год подряд называет специалиста по Data Science «работой №1 в США». Бюро трудовой статистики США сообщает, что к 2026 году ожидается рост занятости в этой сфере на 27.9%. Согласно отчёту MHR Analytics, 80% британских компаний планируют нанять в этом году специалиста по работе с данными или обратиться за консультацией по Data Science. Однако есть не только огромный спрос, но и заметная нехватка квалифицированных аналитиков данных. Всё вышеперечисленное факты однозначно определяют, что Data Science будет востребован ближайшие 20-30 лет, что касается его участия в бизнесе, то иного направления для развития бизнеса нет.

В рамках данной научно-исследовательской работы были рассмотрены: основные понятия Data Science, его математический аппарат, применение в финансах, а также перспектива развития направления в целом. Было проанализировано, какие составляющие в данной науке.

Использование Data Science может существенно улучшить эффективность работы бизнесов и мирового рынка. В процессе работы были пройдены

следующие этапы: 1) Изучение тематических теоретических материалов; 2) Анализ полученной информации; 3) Синтез и изложение полученной информации. Поставленную цель достигнута.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

- 1) Porter, Michael. How Information Gives You Competitive Advantage (англ.) // Harvard Business Review: magazine. Архивировано 23 июня 2015 года.
- 2) INGVARR NET.RU –URL: <http://ingvarr.net.ru/otvet/31-1-0-64443> (дата обращения 07.06.2023);
- 3) Фоксфорд: информационные процессы –URL: <https://foxford.ru/wiki/informatika/informacionnye-processy> (дата обращения 07.06.2023);
- 4) RSF: Найти и обработать. Как наш мозг отбирает и хранит информацию? –URL: https://www.rscf.ru/en/news/media/nayti_i_obrabotat_kak_nash_mozg_otbiraet_i_khranit_informatsiyu (дата обращения 07.06.2023);
- 5) Newneuro: Факты о головном мозге. Интересные факты и мозге для детей. –URL: <https://newneuro.ru/faktyi-o-mozge/#i> (дата обращения 07.06.2023);
- 6) Habr: Механизмы и принципы работы памяти головного мозга человека. –URL: <https://habr.com/ru/articles/371661> (дата обращения 07.06.2023);
- 7) Newneuro: Из чего состоит человеческая память. Принципы работы памяти. –URL: <https://newneuro.ru/chelovecheskaya-pamyat/#i> (дата обращения 07.06.2023);
- 8) Habr: Насколько велик интернет? –URL: <https://habr.com/ru/companies/asus/articles/368853> (дата обращения 07.06.2023);

- 9) Макконнелл К. Р., Брю С. Л. Экономикс: принципы, проблемы и политика : [рус.] = Economics: Principles, Problems, and Policies. — М. : Республика, 1992. — Т. 1. — С. 73—75. — ISBN 5-250-01486-0.
- 10) Mike Lukides. What is data science? The future belongs to the companies and people that turn data into products (англ.). Radar. O'Reilly Media (2 июня 2010). Дата обращения: 2 января 2013. Архивировано 26 января 2013 года.
- 11) Skillbox: Data Science – что это такое. –URL: https://skillbox.ru/media/code/chto_takoe_data_science_i_kto_takoy_data_scientist (дата обращения 07.06.2023);
- 12) Habr: Работа Data Science: что важно знать и как этому научиться. – URL: <https://habr.com/ru/articles/668428> (дата обращения 07.06.2023);
- 13) proglib: Data Science – что это такое простыми словами. –URL: <https://proglib.io/p/chto-takoe-data-science-obyasnyam-prostymi-slovami-2022-12-14> (дата обращения 07.06.2023);
- 14) proglib: Обучение Data Science: какие знания по математике нужны специалисту по анализу данных? –URL: <https://proglib.io/p/obuchenie-data-science-kakie-znaniya-po-matematike-nuzhny-specialistu-po-analizu-dannyh-2020-12-24> (дата обращения 07.06.2023);
- 15) Инфоурок: Краткий конспект к уроку “Информация и данные”. – URL: <https://infourok.ru/kratkiy-konspekt-k-uroku-informaciya-i-dannie-2918730.html> (дата обращения 07.06.2023);
- 16) Инфоурок: Краткий конспект к уроку “Информация и данные”. – URL: <https://infourok.ru/kratkiy-konspekt-k-uroku-informaciya-i-dannie-2918730.html> (дата обращения 07.06.2023);
- 17) РБК тренды: Big Data <https://trends.rbc.ru/trends/innovation/5d6c020b9a7947a740fea65c#:~:text=Big%20Data%20или%20большие%20данные,росте%20объемов%20информации%20в%20мире> (дата обращения 07.06.2023);

- 18) Habr: Подробная история развития облачных технологий. –URL: <https://habr.com/ru/companies/serverspace/articles/730080> (дата обращения 07.06.2023);
- 19) Amazon: История котировок –URL: <https://www.calc.ru/AMZN-istoriya-kotirovok-aktsiy.html> (дата обращения 07.06.2023);
- 20) РЭШ: Зачем дата-сайнтисту экономика? –URL: <https://news.nes.ru/news/zachem-data-sajentistu-ekonomika-chast-1:-sergej-laktionov/> (дата обращения 07.06.2023);
- 21) РЭШ: Зачем дата-сайнтисту экономика? –URL: <https://news.nes.ru/news/zachem-data-sajentistu-ekonomika-chast-1:-sergej-laktionov/> (дата обращения 07.06.2023);
- 22) Datascient: Анализ данных. –URL: <http://datascientist.one/analiz-dannyx-socialnyx-setej-podxody-i-metody> (дата обращения 07.06.2023).

Практикант Ердоган Дениз Ердалович _____
(ФИО полностью) (подпись)

Отчёт о прохождении практической подготовки сдан « ____ » _____ 2023г.

Оценка _____

Руководитель практической
подготовки от Университета Бекмурзаев В.А., к.т.н, доц. _____
(ФИО полностью, должность) (подпись)