МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ

ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

(ФГБОУ ВО «ВГТУ», ВГТУ)

ФАКУЛЬТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И КОМПЬЮТЕРНОЙ БЕЗОПАСНОСТИ

КАФЕДРА КОМПЬЮТЕРНЫХ ИНТЕЛЛЕКТУАЛЬНЫХ ТЕХНОЛОГИЙ ПРОЕКТИРОВАНИЯ

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №1

По дисциплине: «Системы хранения и обработки данных»

Тема: «Сравнительный анализ профессий в области искусственного интеллекта»

Выполнил работу студент группы мИИВТ-231 Романов Д.А.

Принял: Короленко В.В.

Защищена \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Оценка \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Воронеж 2023

**Цели работы:**

изучение основных профессий в области искусственного интеллекта.

**Основные задачи:**

* описание профессии «Аналитик данных (Data Scientist)»;

**Ход работы:**

В данной работе была разобрана одна из востребованных профессий современности – аналитик данных. На основе полученной информации была сделана презентация с описанием особенностей профессии.

Аналитик данных представляет собой специалиста, который занимается сбором, анализом и интерпретацией больших объемов данных с целью выявления тенденций, разработки прогнозов, выявления закономерностей и обеспечения поддержки принятия решений на основе данных. Данная информация отображена на слайде презентации (Рисунок 1).

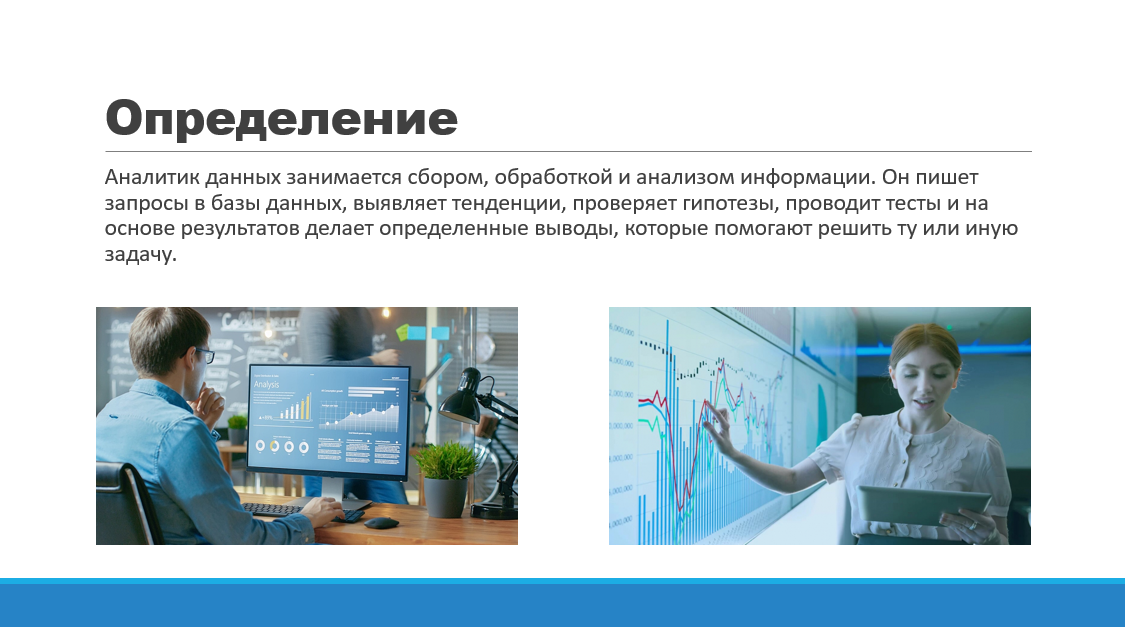


Рисунок 1 – Определение аналитика данных

Типовые задачи, решаемые аналитиком данных, могут различаться в зависимости от места работы и уровня квалификации. Базовый список включает в себя следующие задачи:

* Прогнозирование;
* Оптимизация производственных процессов;
* Визуализация данных и создание отчетов;
* Разработка стратегий маркетинга и продаж;
* Идентификация тенденций и паттернов.

В презентации задачи аналитика данных представлены на слайде 2-3 (Рисунок 2 – 3).



Рисунок 2 – Задачи аналитика



Рисунок 3 – Задачи аналитика

Далее были рассмотрены требования к специалистам, необходимые знания и навыки. Образованиетребуется высшее образование в области математики, статистики, экономики, компьютерных наук или смежной области. Из технических навыков отмечают умение программировать на Python, R, SQL, понимание баз данных и обработка данных, английский уровня Advanced Proficiency для чтения технической литературы, понимание специфики бизнеса и доменной области. Также выделяют важными такие навыки, как умение презентовать результаты работы и умение коммуницировать с коллегами. Данные требования также были отображены в презентации (Рисунок 4).

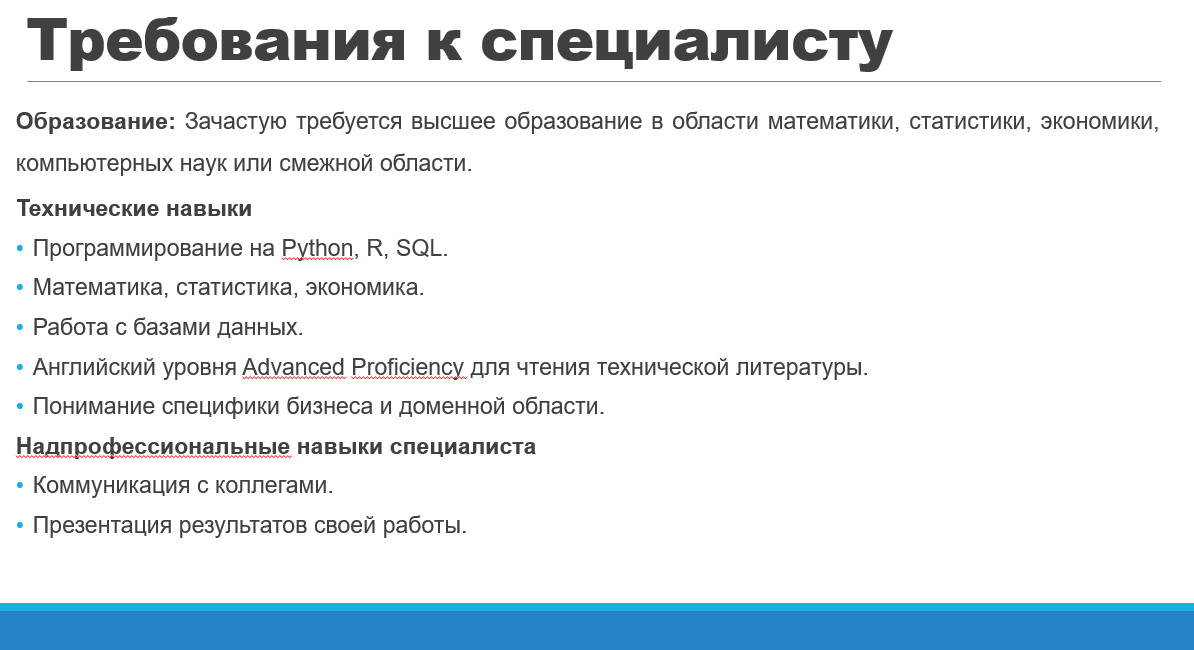


Рисунок 4 – Определение аналитика данных

Были определены средства и инструменты, используемые специалистом в профессиональной деятельности, наиболее популярные из них были изображены в презентации (рисунок 5). Например, самыми используемыми языками программирования являются python, SQL и R.

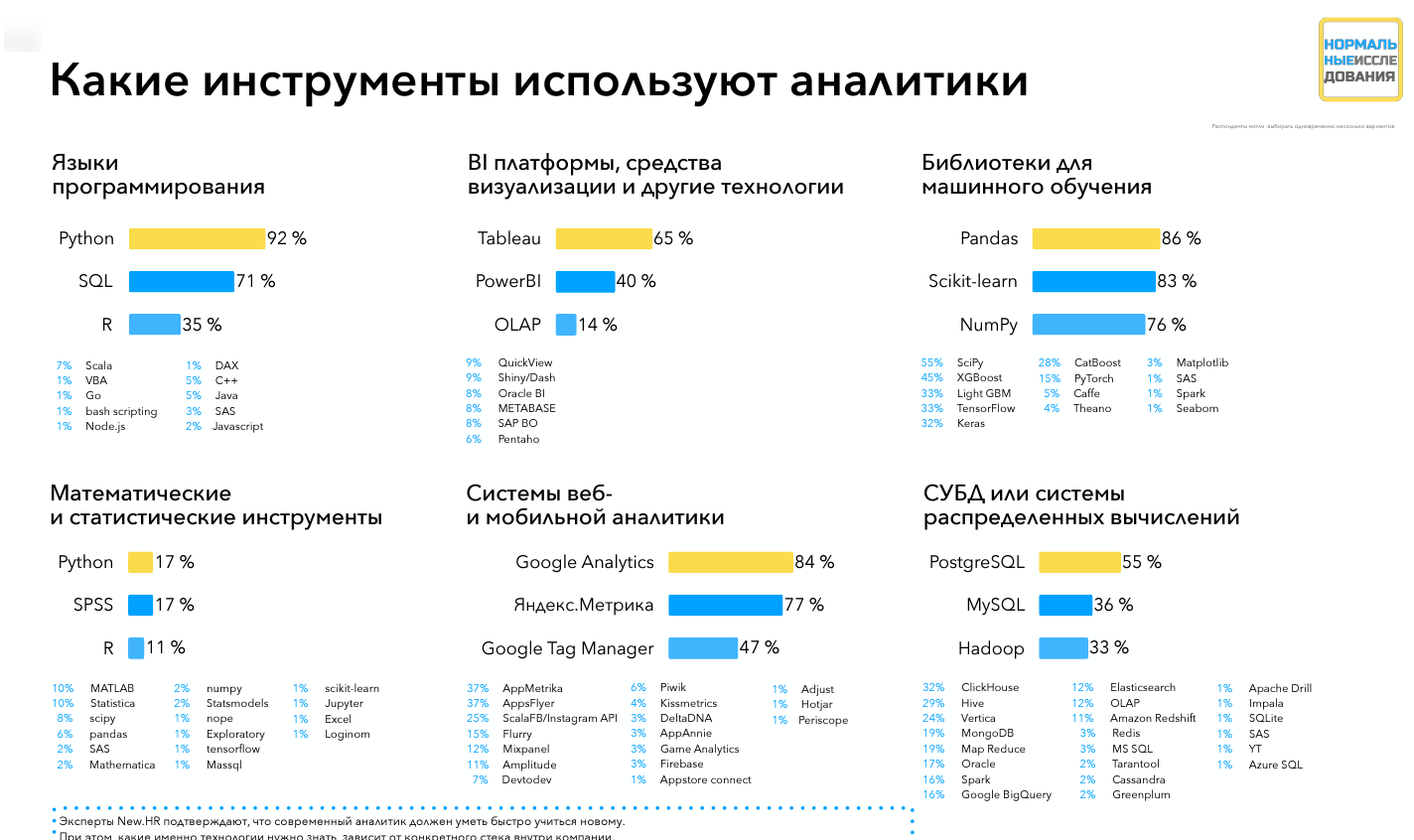


Рисунок 5 – Средства специалиста

В команде аналитик данных играет важную роль. Он отвечает за анализ больших объемов информации, выявляет тенденции, прогнозирует результаты и помогает в принятии решений на основе данных. Работает аналитик с инженерами по обработке данных, специалистами по машинному обучению и бизнес-аналитики. Аналитик данных помогает в разработке стратегий по работе с данными, анализирует данные и презентует результаты анализа коллегам, принимает решения, основываясь на своих результатах. Данная информация изображена в презентации (Рисунок 6).

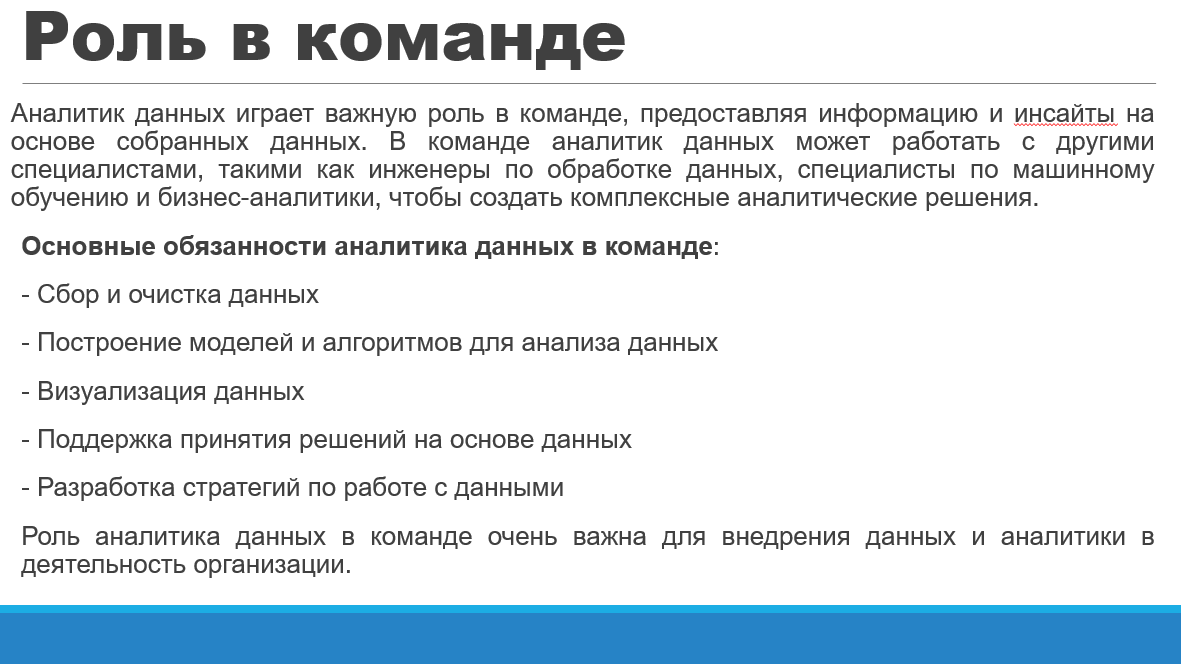


Рисунок 6 – Роль аналитика в команде

Также стоит отметить, что аналитик данных имеет множество пересечений со специалистом по Data Science. Однако, это две разные профессии. Отличие заключается в том, что в отличие от аналитика данных, специалист по Data Science обычно более технически подкован и имеет глубокие знания в области программирования, математической статистики и машинного обучения, в то время как для аналитика данных основной задачей является интерпретация результатов и принятие бизнес-решений. В презентации о различии двух специалистов написано на 7-м слайде (Рисунок 7).



Рисунок 7 – Анализ данных и Data Science

Важным вопросом любой профессии является вопрос о ее актуальности и востребованности. Касательно аналитиков данных, спрос на них высок и продолжает расти. Также квалифицированные специалисты могут рассчитывать на высокую зарплату и перспективы для роста. О востребованности данной профессии был создан последний слайд презентации (Рисунок 8).

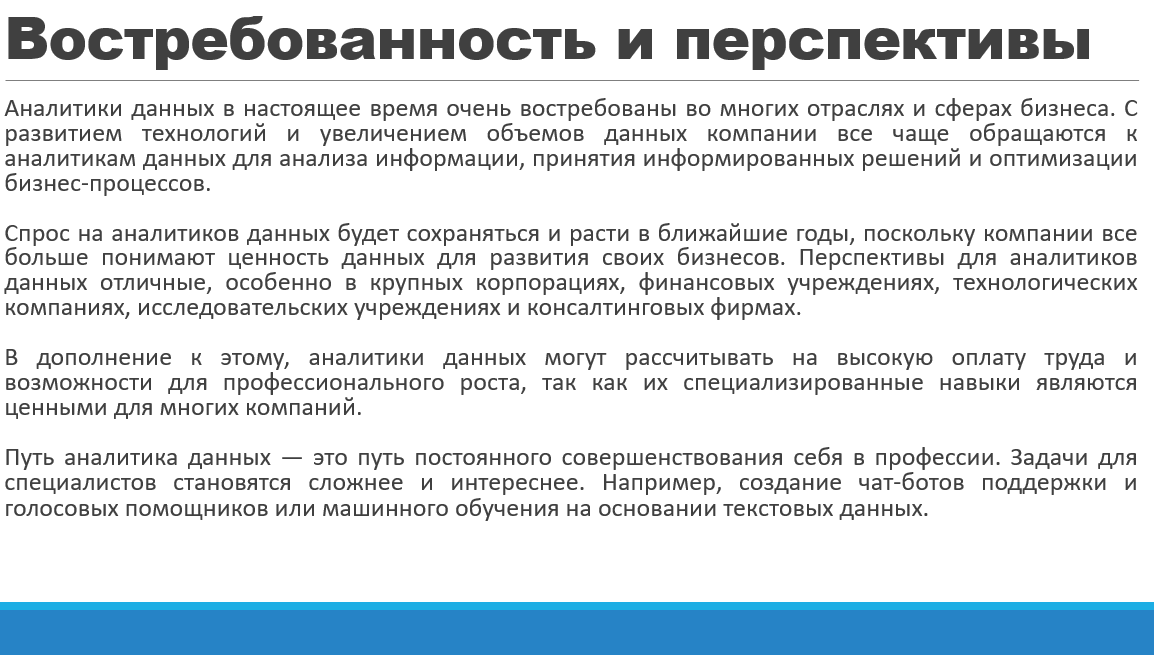


Рисунок 8 – Востребованность и перспективы

**Вывод**: в данной лабораторной работе были изучены основные профессии в области искусственного интеллекта, в частности профессия «Аналитик данных».

**Контрольные вопросы**

1. Дайте определение термина «Искусственный интеллект».

Искусственный интеллект – технология, позволяющая системе, машине или компьютеру выполнять задачи, требующие разумного мышления, то есть имитировать поведение человека для постепенного обучения с использованием полученной информации и решения конкретных вопросов.

1. Дайте определение термина «Наука о данных».

Наука о данных – профессиональная деятельность, связанная с эффективным и максимально достоверным поиском закономерностей в данных, извлечение знаний из данных в обобщённой форме, а также их оформление в виде, пригодном для обработки заинтересованными сторонами (людьми, программными системами, управляющими устройствами) в целях принятия обоснованных решений.

1. Дайте определение термина «Машинное обучение».

Машинное обучение – это методы искусственного интеллекта, которые позволяют построить обучаемые модели для разных целей: например, автоматизации процессов, автоматического перевода текстов, распознавания изображений.

1. Дайте определение термина «Анализ данных».

Анализ данных – это область между математикой и информатикой, которая занимается построением и исследованием методов извлечения полезной информации из данных. Анализ данных позволяет преобразовать массив данных в выводы, на основе которых будут приниматься решения и строиться действия. Он охватывает большое количество методов из разных областей науки.

1. Что означает термин «Big Data» в информационных технологиях?

Big Data – это сложные и объёмные наборы разной информации. Они представлены в «сыром виде» и требуют предварительной обработки, чтобы получить из них ценные сведения, которые могут принести пользу предприятиям и организациям.

1. Какие главные характеристики Big Data?

Объем:

Количество данных – важный фактор. Располагая ими в больших количествах, Вам потребуется обрабатывать большие объемы неструктурированных данных низкой плотности. Ценность таких данных не всегда известна. Это могут быть данные каналов Twitter, данные посещаемости веб-страниц, а также данные мобильных приложений, сетевой трафик, данные датчиков. В некоторые организации могут поступать десятки терабайт данных, в другие — сотни петабайт.

Скорость:

Скорость в данном контексте – это скорость приема данных и, возможно, действий на их основе. Обычно высокоскоростные потоки данных поступают прямо в оперативную память, а не записываются на диск. Некоторые "умные" продукты, функционирующие на основе Интернета, работают в режиме реального или практически реального времени. Соответственно, такие данные требуют оценки и действий в реальном времени.

Разнообразие:

Разнообразие означает, что доступные данные принадлежат к разным типам. Традиционные типы данных структурированы и могут быть сразу сохранены в реляционной базе данных. С появлением Big Data данные стали поступать в неструктурированном виде. Такие неструктурированные и полуструктурированные типы данных как текст, аудио и видео, требуют дополнительной обработки для определения их значения и поддержки метаданных.

1. Соотношение бизнес-аналитики (Business Intelligence) и науки о данных (Data Science).

Data science и BI различаются по конечному результату, полученному характерными для них методами работы с данными. Для BI результирующим продуктом является информация, а для data science таким продуктом служат знания. Общим же является то, что как информация, так и знания извлекаются из данных с участием человека и предназначены для передачи другому человеку.

Второй фактор, определяющий различия – полнота используемых данных. Любые данные, в конечном счете, являются отражением окружающего мира, но эти отражения могут различаться по полноте представления. В BI используются только структурированные цифровые данные, дающие весьма ограниченную картину окружающего мира, а в data science могут быть использованы любые данные, достаточные для отражения картины окружающего мира с любой требуемой полнотой.

1. Что такое жизненный цикл анализа данных?

Жизненный цикл – это последовательность шагов и процессов, которые выполняются при проведении анализа данных от начала до включает в себя следующие этапы:

1. Постановка задач;

2. Сбор данных;

3. Подготовка данных;

4. Анализ данных;

5.Презентация результатов;

6. Применение результатов;

7. Мониторинг и обновление.

Жизненный цикл анализа данных позволяет систематизировать и организовать процесс работы с данными, что помогает создать надежные и точные результаты анализа.