МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ 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 Ахлестин А.И. (подпись) Фамилия, инициалы

Принял \_\_\_\_\_\_\_\_\_\_\_\_Короленко В.В.  
 (подпись) Фамилия, инициалы

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

Воронеж 2023

Цель работы: изучение основных профессий в области искусственного интеллекта.

Задание на лабораторную работу: необходимо описать профессию «Инженер данных», обратить внимание на следующие аспекты:

1. типовые задачи, решаемые специалистом;
2. требуемое образование, необходимые знания и навыки;
3. средства, используемые специалистом в профессиональной деятельности (оборудование, программное обеспечение, методические материалы, информационные ресурсы);
4. команда, в которой работает соответствующий специалист, роль в команде, описание взаимодействия с членами команды;
5. соотношение со смежными профессиями;
6. перспективы профессионального роста для соответствующего специалиста.

Ход выполнения задания:

Профессия "Инженер данных" является одной из наиболее востребованных и перспективных в области информационных технологий. Инженер данных занимается сбором, обработкой, анализом и интерпретацией данных с целью предоставления бизнесу ценных инсайтов.

# 1 Типовые задачи, решаемые специалистом

# Инженер данных решает разнообразные задачи, связанные с обработкой и анализом данных. Он занимается сбором, хранением, и подготовкой данных для дальнейшего использования в бизнес-процессах и аналитике. Исходя из этого, типовые задачи, которые решает инженер данных, включают:

# Создание и обслуживание баз данных: инженер данных разрабатывает и поддерживает структуры данных, необходимые для хранения информации.

# Сбор данных: специалист собирает данные из различных источников, таких как базы данных, веб-сервисы, внешние API и датчики.

# Очистка и трансформация данных: инженер данных обрабатывает и структурирует данные, устраняет ошибки и выполняет преобразования, чтобы данные стали пригодными для анализа.

# Разработка ETL-процессов: он создает и оптимизирует процессы извлечения, преобразования и загрузки данных, которые позволяют переносить данные из источников в хранилище данных.

# Работа с большими данными: инженеры данных могут заниматься обработкой и анализом больших объемов данных, используя специализированные технологии, такие как Apache Hadoop и Apache Spark.

# Обеспечение безопасности данных: инженеры данных также отвечают за защиту данных и соблюдение нормативных требований в области безопасности.

# Разработка и оптимизация инфраструктуры данных: они создают и поддерживают инфраструктуру для хранения и доступа к данным, обеспечивая масштабируемость и производительность.

# Сотрудничество с аналитиками данных и бизнес-пользователями: инженеры данных работают в команде с аналитиками и предоставляют им доступ к данным, необходимым для принятия стратегических решений.

# Визуализация данных: они могут также создавать визуализации и отчеты, чтобы представить данные более наглядно для бизнес-пользователей.

# Разработка и внедрение решений для машинного обучения: В некоторых случаях инженеры данных работают с моделями машинного обучения и помогают внедрять их в бизнес-процессы.

# 2 Требуемое образование, необходимые знания и навыки

# Для занятия профессией инженера данных требуется бакалаврский или магистерский диплом в области компьютерных наук, информатики, статистики или смежной области. Это образование предоставляет теоретические знания и фундаментальную базу для работы с данными.

# Кроме того, инженер данных должен обладать знанием языков программирования, таких как SQL, Python или Java, а также иметь глубокое понимание статистики и методов анализа данных. Это помогает ему разрабатывать и оптимизировать процессы обработки данных и создавать алгоритмы для анализа информации.

# Инженер данных также должен владеть навыками работы с базами данных и инструментами ETL (извлечения, преобразования и загрузки данных), что позволяет ему создавать эффективные процессы обработки информации. Опыт работы с системами управления версиями и умение визуализировать данные являются важными аспектами его профессионального навыка.

# Коммуникационные навыки также не менее важны, так как инженер данных часто взаимодействует с аналитиками, разработчиками и бизнес-пользователями. Умение объяснить сложные технические концепции непрофессионалам и работать в команде являются ключевыми аспектами успешной карьеры в этой профессии.

# 3 Средства, используемые специалистом

# Инженер данных использует различные средства и инструменты в своей профессиональной деятельности. Он оперирует программным обеспечением, таким как базы данных (например, MySQL, PostgreSQL), инструменты для ETL (например, Apache NiFi, Talend), а также инструменты для анализа данных (например, Jupyter, Tableau).

# Для работы с большими данными инженеры данных могут использовать специализированные технологии, такие как Apache Hadoop и Apache Spark. Они также могут работать с оборудованием, таким как высокопроизводительные серверы и кластеры, чтобы обеспечить эффективную обработку и хранение данных.

# Средства и оборудование, выбираемые инженером данных, зависят от конкретных требований проекта и объема данных, с которыми он работает. Умение выбирать и настраивать подходящие средства для каждой конкретной задачи является важным аспектом его работы.

# 4 Роль в команде

# Инженер данных работает в командной среде, где в зависимости от проекта и организации его роль может варьироваться. В команде он часто взаимодействует с аналитиками данных, научными сотрудниками, разработчиками и другими профессионалами. Его роль в команде заключается в создании и поддержании инфраструктуры данных, обеспечении доступа к данным для аналитиков и научных сотрудников, а также сотрудничестве с бизнес-аналитиками и руководством компании для определения бизнес-требований и целей анализа данных.

# Инженер данных также может выступать в качестве посредника между техническими и бизнес-командами, помогая перевести бизнес-запросы в технические решения и наоборот. Его способность коммуникации и понимание как технических, так и бизнес-аспектов проекта позволяют ему содействовать успешной реализации аналитических задач и достижению бизнес-целей компании.

# 5 Соотношение со смежными профессиями

# Инженер данных имеет смежные профессии, с которыми он взаимодействует и сотрудничает в рамках проектов. Эти смежные профессии включают в себя аналитиков данных, разработчиков, администраторов баз данных и научных сотрудников.

# С аналитиками данных инженер данных совместно работает над пониманием требований анализа данных, созданием структурированных наборов данных и обеспечением им доступа к данным. Они совместно разрабатывают процессы извлечения и трансформации данных (ETL) и обсуждают стратегии анализа.

# С разработчиками инженер данных может сотрудничать в создании приложений и систем, которые взаимодействуют с данными, а также при интеграции данных в прикладные программы.

# С администраторами баз данных он взаимодействует в случае необходимости управления и обслуживания баз данных, где хранятся данные.

# С научными сотрудниками инженер данных может сотрудничать в разработке и внедрении моделей машинного обучения и анализе данных в научных исследованиях.

# Сотрудничество с этими смежными профессиями позволяет инженеру данных обеспечивать успешное создание и использование данных в различных бизнес- и научных контекстах.

# 6 Перспективы профессионального роста

# Инженер данных имеет хорошие перспективы профессионального роста. С накоплением опыта и навыков он может продвигаться по карьерной лестнице и занимать более ответственные позиции, такие как старший инженер данных, архитектор данных или руководитель проектов в области данных.

# Помимо этого, инженер данных может выбрать специализацию в определенных областях, таких как инженерия машинного обучения или аналитика данных, что позволит ему развивать глубокие знания и экспертизу в конкретной области.

# Такие специализации могут открыть возможности для работы в более специализированных проектах и компаниях, где требуется углубленное понимание конкретных аспектов обработки и анализа данных.

# Благодаря быстрому развитию технологий и постоянному увеличению объемов данных в современном мире, специалисты по данным становятся все более востребованными, что создает много перспектив для развития карьеры в этой области.

Вывод: В данной лабораторной работе была представлена обширная информация о профессии "Инженер данных". Мы рассмотрели типовые задачи, которые решает этот специалист, а также обсудили требуемое образование, необходимые знания и навыки. Описаны средства, которые инженер данных использует в своей работе, и его роль в команде, включая взаимодействие с другими специалистами. Также рассмотрено соотношение с смежными профессиями и перспективы профессионального роста.

Контрольные вопросы:

1. Искусственный интеллект:

Искусственный интеллект (ИИ) представляет собой область информатики, занимающуюся созданием систем, способных выполнять задачи, обычно требующие интеллектуальных способностей человека. Эти системы могут включать в себя машинное обучение, обработку естественного языка, распознавание образов, планирование, принятие решений и другие аспекты.

1. Наука о данных:

Наука о данных (Data Science) - это междисциплинарная область знаний, которая использует методы, процессы, алгоритмы и системы для извлечения знаний и информации из структурированных и неструктурированных данных. Она объединяет статистику, анализ данных, машинное обучение и другие техники для понимания и анализа данных.

1. Машинное обучение.

Машинное обучение (Machine Learning) - это подраздел искусственного интеллекта, который обучает компьютерные системы выполнять задачи, не явно программированные для них. Оно позволяет системам самостоятельно обучаться на основе опыта и данных, делать прогнозы, распознавать образы и принимать решения без явного программирования.

1. Анализ данных.

Анализ данных - это процесс обработки, изучения и интерпретации данных с целью выявления полезной информации, выявления закономерностей, выявления тенденций и поддержки принятия решений.

1. Big Data.

Big Data (большие данные) относится к объемам данных, которые находятся за пределами возможностей традиционных баз данных и инструментов обработки данных. Характеризуется тремя основными аспектами - объемом (большие объемы данных), скоростью (высокая скорость поступления данных) и разнообразием (различные типы данных).

1. Характеристики Big Data.

Объем (Volume), Скорость (Velocity), Разнообразие (Variety), Достоверность (Veracity), Вариабельность (Volatility), Величина (Value). Эти характеристики описывают особенности больших объемов и разнообразия данных, требующих новых методов и технологий для их обработки.

1. Соотношение бизнес-аналитики и науки о данных.

Бизнес-аналитика (Business Intelligence) и наука о данных (Data Science) взаимосвязаны, но имеют различные цели. Бизнес-аналитика фокусируется на анализе и визуализации данных для поддержки бизнес-решений, тогда как наука о данных включает в себя более широкий спектр методов, включая машинное обучение и алгоритмы для извлечения знаний из данных.

1. Жизненный цикл анализа данных.

Жизненный цикл анализа данных представляет собой последовательность этапов, которые проходят данные от сбора до анализа и использования результатов. Этапы могут включать в себя постановку задач, сбор данных, очистку данных, исследовательский анализ, построение моделей, валидацию и интерпретацию результатов, а также представление результатов бизнес-пользователям.