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

ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №1

По дисциплине: «Системы хранения и обработки данных»

Тема: «Сравнительный анализ профессий в области искусственного интеллекта»

Выполнил работу студент группы мИИВТ-241: Черкасов Н.Е.

подпись, дата

Принял: Короленко В.В.

подпись, дата

Воронеж 2024

Задачи лабораторной работы:

1) Необходимо описать профессию, обратить внимание на следующие аспекты:

- типовые задачи, решаемые специалистом;

- требуемое образование, необходимые знания и навыки;

- средства, используемые специалистом в профессиональной деятельности (оборудование, программное обеспечение, методические материалы, информационные ресурсы);

- команда, в которой работает соответствующий специалист, роль в команде, описание взаимодействия с членами команды;

- соотношение со смежными профессиями;

перспективы профессионального роста для соответствующего специалиста.

2) Подготовить отчёт о проделанной работе и доклад.

1. ML-инженер.

**Типовые задачи, решаемые специалистом:**

ML Engineer (инженер по машинному обучению) занимается разработкой, внедрением и оптимизацией моделей машинного обучения. Основные задачи включают:

1) Разработка и обучение моделей машинного обучения для решения конкретных задач, таких как классификация, регрессия, кластеризация и прогнозирование.

2) Оптимизация и настройка гиперпараметров моделей для повышения их точности и производительности.

3) Интеграция моделей машинного обучения в производственные системы и приложения.

4) Обработка и анализ больших объемов данных для подготовки их к обучению моделей.

5) Создание и поддержка инфраструктуры для обучения и развертывания моделей (например, использование облачных платформ).

6) Мониторинг и обновление моделей в реальном времени для обеспечения их актуальности и точности.

**Требуемое образование, необходимые знания и навыки:**

- Высшее образование в области компьютерных наук, математики, статистики или смежных дисциплин.

- Глубокие знания в области машинного обучения, включая алгоритмы, такие как линейная регрессия, деревья решений, нейронные сети и ансамбли моделей.

- Опыт работы с языками программирования, такими как Python, R, Java или C++.

- Знание библиотек и фреймворков для машинного обучения, таких как TensorFlow, PyTorch, Scikit-learn, Keras.

- Понимание методов обработки данных (Data Preprocessing) и работы с большими данными (Big Data).

- Навыки работы с облачными платформами (AWS, Google Cloud, Azure) и инструментами для развертывания моделей (Docker, Kubernetes).

- Умение работать с базами данных (SQL, NoSQL) и системами управления данными.

**Средства, используемые специалистом в профессиональной деятельности:**

1) Оборудование: мощные серверы с GPU/TPU для обучения моделей, облачные платформы для хранения и обработки данных.

2) Программное обеспечение:

- Языки программирования: Python, R, Java.

- Фреймворки для машинного обучения: TensorFlow, PyTorch, Scikit-learn.

- Инструменты для работы с данными: Pandas, NumPy, Spark.

- Облачные платформы: AWS SageMaker, Google AI Platform, Azure ML.

- Инструменты для развертывания моделей: Docker, Kubernetes, Flask.

3) Методические материалы: учебники по машинному обучению, научные статьи, онлайн-курсы (Coursera, edX, Udemy).

4) Информационные ресурсы: научные журналы, блоги, форумы (Kaggle, GitHub, Stack Overflow).

**Команда, в которой работает специалист:**

ML Engineer взаимодействует с различными специалистами, включая:

- Data Scientists — для разработки и тестирования моделей.

- Data Engineers — для подготовки и обработки данных.

- Программисты и разработчики — для интеграции моделей в приложения

- DevOps-инженеры — для развертывания и мониторинга моделей в производственной среде.

**Соотношение со смежными профессиями:**

ML Engineer тесно связан с такими профессиями, как Data Scientist, Data Engineer, Software Developer и DevOps Engineer.

Data Scientist занимается анализом данных и созданием моделей, но не всегда отвечает за их внедрение в производство.

- Data Engineer фокусируется на подготовке и обработке данных, создании инфраструктуры для их хранения и передачи.

- Software Developer занимается разработкой программного обеспечения, но не всегда работает с моделями машинного обучения.

**Перспективы профессионального роста:**

Специалисты в области машинного обучения могут развиваться в нескольких направлениях:

- Углубление знаний в области глубокого обучения (Deep Learning) и нейронных сетей.

- Развитие в направлении архитектуры машинного обучения (ML Architect) или специалиста по искусственному интеллекту (AI Specialist).

- Участие в исследовательских проектах и разработка новых алгоритмов машинного обучения.

- Переход в смежные области, такие как Data Science, Big Data или Cloud Engineering.

Таким образом, профессия ML Engineer является одной из самых востребованных в современном мире технологий, открывающей широкие возможности для карьерного роста и профессионального развития.

Вывод: при выполнении лабораторной работы была описана профессия ML Engineer. Были рассмотрены типовые задачи, решаемые специалистом, необходимые знания и навыки, используемые средства, а также перспективы профессионального роста.

Контрольные вопросы:

1) Искусственный интеллект (ИИ) — это область информатики, занимающаяся созданием систем, способных выполнять задачи, требующие интеллекта и человеческих навыков, таких как понимание языка, распознавание образов, принятие решений и обучение. ИИ включает в себя различные подходы, такие как правила, машинное обучение и нейронные сети.

2) Наука о данных — это междисциплинарная область, занимающаяся извлечением знаний из данных с использованием методов статистики, математики, информатики и других наук. Основная цель науки о данных — выявление закономерностей, создание прогнозирующих моделей и принятие обоснованных решений на основе анализа данных.

3) Машинное обучение — это класс методов искусственного интеллекта, которое позволяет системам автоматически обучаться на основе данных и улучшать результаты своих действий без явного программирования. Машинное обучение включает в себя методы, такие как supervised learning (обучение с учителем), unsupervised learning (обучение без учителя) и reinforcement learning (обучение с подкреплением).

4) Анализ данных — это процесс очистки, преобразования и моделирования данных с целью получения полезной информации для принятия решений. Анализ данных может включать в себя описательную статистику, визуализацию данных и использование алгоритмов машинного обучения для извлечения инсайтов.

5) Big Data — это термин, описывающий огромные объемы данных, которые сложно обрабатывать с помощью традиционных методов и инструментов. Big Data включает в себя разнообразные типы данных из различных источников и требует специальных технологий для сбора, хранения, обработки и анализа.

6) Главные характеристики Big Data:

- Объем (Volume) — огромные объемы данных, которые растут экспоненциально.

- Разнообразие (Variety) — данные могут иметь различные форматы (структурированные, полуструктурированные, неструктурированные).

- Скорость (Velocity) — данные создаются и обновляются в реальном времени, что требует быстрого анализа.

- Истинность (Veracity) — учет качества и достоверности данных при их анализе.

- Ценность (Value) — извлечение полезной информации из больших массивов данных.

7) Соотношение бизнес-аналитики (Business Intelligence) и науки о данных (Data Science): Бизнес-аналитика в первую очередь фокусируется на обработке и анализе исторических данных для поддержки бизнес-решений и оптимизации процессов. Наука о данных, в свою очередь, использует более сложные методы, включая машинное обучение и статистику, для моделирования и прогнозирования, часто ориентируясь на создание новых инсайтов и инновационных решений. В то время как Business Intelligence может включать в себя основы анализа данных, Data Science предоставляет более широкий набор инструментов и подходов.

8) Жизненный цикл анализа данных включает несколько этапов, таких как:

- Определение проблемы и цели анализа.

- Сбор данных из различных источников.

- Очистка и подготовка данных для анализа.

- Анализ данных с использованием различных методов.

- Интерпретация и представление результатов.

- Применение инсайтов для принятия решений и действий.

- Мониторинг и обновление моделей в зависимости от новых данных.

Таким образом, все эти термины и концепции имеют ключевое значение в современном мире технологий и бизнеса.