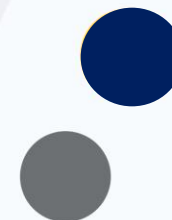


Интеллектуальные информационные системы

Лекция 7. Основные роли в Machine Learning проектах

Кафедра информатики
Институт кибербезопасности и цифровых технологий
РТУ МИРЭА



Основные роли в Machine Learning проектах

• • Data Scientist

Специалист, который создает ML модели и проводит глубокий анализ данных для решения бизнес-задач с помощью методов машинного обучения и статистики.

Основные обязанности:

- Исследование и анализ данных для выявления паттернов и зависимостей
- Разработка и оптимизация ML моделей под конкретные бизнес-задачи
- Оценка метрик и производительности моделей
- Экспериментирование с различными алгоритмами и подходами
- Документирование методологии и результатов исследований

Технологический стек:

- Языки: Python (основной), R (статистический анализ), SQL
- ML фреймворки: TensorFlow/Keras, PyTorch, Scikit-learn
- Анализ данных: Pandas, NumPy, Jupyter Notebooks
- Визуализация: Matplotlib, Seaborn
- Эксперименты: Weights & Biases, MLflow

Основные роли в Machine Learning проектах

ML Engineer

Специалист, отвечающий за внедрение ML моделей в производство и создание эффективной инфраструктуры для их работы.

Основные обязанности:

- Перенос моделей в производственную среду (ML Ops)
- Оптимизация моделей для промышленного использования
- Создание и поддержка пайплайнов для обработки данных
- Мониторинг производительности моделей в production
- Интеграция ML-компонентов с общей инфраструктурой

Технологический стек:

- Языки: Python, C++ (оптимизация), Java
- ML Ops: MLflow, Kubeflow, TensorFlow Serving
- API: FastAPI, Flask
- Распределенные вычисления: Ray, Spark ML
- Оркестрация: Airflow

Основные роли в Machine Learning проектах

• Data Engineer

Специалист по созданию и поддержке инфраструктуры для работы с данными, обеспечивающий их качественную обработку, хранение и доступность.

Основные обязанности:

- Создание и поддержка data pipeline
- Обеспечение качества и доступности данных
- Проектирование систем хранения данных
- Оптимизация процессов обработки данных
- Интеграция с различными источниками данных

Технологический стек:

- Языки: Python, SQL, Scala
- Big Data: Apache Spark, Hadoop
- Потокковая обработка: Apache Kafka
- Оркестрация: Airflow, Apache Beam
- Базы данных: PostgreSQL, MongoDB
- Data Build Tool: dbt

Основные роли в Machine Learning проектах

DevOps Engineer

Специалист, обеспечивающий бесперебойную работу инфраструктуры и автоматизацию процессов разработки и развертывания.

Основные обязанности:

- Настройка и поддержка инфраструктуры
- Автоматизация процессов разработки и деплоя
- Мониторинг системных ресурсов
- Обеспечение безопасности и отказоустойчивости
- Управление контейнеризацией и оркестрацией

Технологический стек:

- Языки: Python, Shell scripting (Bash), Go, YAML
- Контейнеризация: Docker, Kubernetes
- CI/CD: Jenkins, GitLab CI
- IaC: Terraform, Ansible
- Мониторинг: Prometheus, Grafana
- Логирование: ELK Stack

Основные роли в Machine Learning проектах

Data Analyst

Специалист по анализу данных, который превращает сырые данные в ценные бизнес-инсайты и помогает принимать решения на основе данных.

Основные обязанности:

- Сбор и обработка данных из различных источников
- Создание аналитических отчетов и дашбордов
- Проведение A/B тестирований
- Выявление трендов и паттернов в данных
- Формулирование бизнес-рекомендаций на основе анализа
- Построение и поддержка ETL процессов
- Разработка KPI и метрик эффективности

Основные обязанности:

- Языки: Python, SQL, R
- BI инструменты:
 - Tableau
 - Power BI
 - Superset
- Анализ данных:
 - Pandas
 - NumPy
 - Excel/Google Sheets
- Статистические инструменты:
 - SciPy
 - Statsmodels
- Визуализация:
 - Matplotlib
 - Seaborn
 - ggplot2
 - Plotly

Основные роли в Machine Learning проектах

Кто есть кто в data science

<https://education.yandex.ru/journal/kto-est-kto-v-data-science-analitiki-issledovateli-ml-razrabotchiki-i-inzhenery>

