

Специалист по работе с данными и применению ИИ NLP-инженер

2 года

продолжительность программы

120 зачётных единиц

1 семестр

Основы работы с данными

31 з. е.

- Программирование на Python 5 з. е.
- Предподготовка и анализ данных 6 з. е.
- Машинное обучение с учителем 3 з. е.
- Базы данных и язык SQL 5 з. е.
- BI и методы визуализации данных 3 з. е.
- Специальные разделы теории вероятностей 3 з. е.
- Специальные разделы математической статистики 3 з. е.
- A/B-тестирование 3 з. е.

2 семестр

Изучение Data Science

29 з. е.

- Алгоритмы и структуры данных 7 з. е.
- Машинное обучение без учителя 3 з. е.
- Принципы разработки на Python 5 з. е.
- Операционные системы семейства Unix 4 з. е.
- Учебная практика 2 з. е.
- Прогнозирование временных рядов 5 з. е.
- Навыки коммуникации и представления данных 3 з. е.

3 семестр

Погружение в NLP

29 з. е.

- Работа с изображениями и текстами 3 з. е.
- Классические и глубинные модели по работе с текстом 3 з. е.
- Нейросетевые модели для последовательностей 3 з. е.
- Модели на базе архитектуры трансформер 4 з. е.
- Большие языковые модели 3 з. е.
- Генеративные модели для речи и текста 3 з. е.
- Учебная практика 7 з. е.
- Гибкие методологии 3 з. е.

4 семестр

Подготовка к ВКР

31 з. е.

- Производственная практика (научно-исследовательская работа) 22 з. е.
- Преддипломная практика 3 з. е.
- Выполнение и защита выпускной квалификационной работы 6 з. е.

1 семестр

Программирование на Python

МИФИ

5 з. е.

Научитесь программировать на Python, изучите основные структуры данных, принципы ООП и обработку исключений. Поймёте, как писать чистый и эффективный код, работать с библиотеками для анализа данных. Разберётесь в основах тестирования и инструментах контроля версий.

Примеры проектов

Создадите программу для регистрации и отслеживания продуктов в холодильнике, а также реализуете классическую игру «Змейка» на Python.

Инструменты

Git, GitHub, VS Code, Flake8, Pytest, bash, venv.

Предподготовка и анализ данных

Практикум

6 з. е.

Научитесь очищать данные от выбросов, пропусков и дубликатов, а также преобразовывать разные форматы данных.

Примеры проектов

Проанализируете данные о клиентах банка и определите, какие показатели влияют на их кредитоспособность.

Инструменты

Python, Pandas.

Машинное обучение с учителем

МИФИ

3 з. е.

Познакомитесь с библиотекой scikit-learn и циклом создания модели машинного обучения. Узнаете, как подготовить данные для машинного обучения и проверить работу модели. Освоите линейные и нелинейные модели. Научитесь решать задачи регрессии и классификации. Освоите новые метрики качества модели и подходы для их улучшения. Изучите основы линейной алгебры.

Примеры проектов

Разработаете модель машинного обучения, которая поможет владельцам молочной фермы принимать эффективные решения.

Инструменты

Scikit-learn, SVM, kNN, NumPy.

Базы данных и язык SQL

МИФИ

5 з. е.

Разберётесь, что такое база данных, и научитесь писать сложные запросы к ней. Поймёте, как считать ключевые бизнес-метрики через SQL-запросы и подключаться к базам данных из Python.

Примеры проектов

Напишете ряд запросов разной сложности к базе данных, в которой хранится информация о венчурных инвесторах, стартапах и инвестициях в них. С помощью Python и SQL подключитесь к базе данных, посчитаете и визуализируете ключевые метрики сервис-системы вопросов и ответов о программировании.

Инструменты

SQL, PostgreSQL.

BI и методы визуализации данных

Практикум

3 з. е.

Познакомьтесь с принципами бизнес-аналитики и научитесь превращать данные в ценные бизнес-решения. Освойте методы сбора, обработки и визуализации данных. Изучите основные метрики, показатели эффективности и способы их интерпретации. Будете работать с инструментами BI, строить дашборды и автоматизировать отчётность.

Примеры проектов

Создадите интерактивный дашборд для анализа продаж и клиентского поведения. Соберёте данные из базы, подготовите их для анализа и визуализируете ключевые метрики. Настроите автоматическое обновление отчёта и создадите удобный интерфейс с дашбордом в BI-системе. Проанализируете тренды, выявите аномалии и сформируете рекомендации для бизнеса.

Инструменты

DataLens, Superset.

Специальные разделы теории вероятностей

МИФИ

3 з. е.

Познакомьтесь с основами теории вероятностей и их применением в анализе данных. Изучите законы распределения случайных величин, методы оценки вероятностей и основные теоремы. Научитесь работать с дискретными и непрерывными распределениями, вычислять математическое ожидание, дисперсию и ковариацию. Разберётесь в методах байесовского анализа, законах больших чисел и центральной предельной теореме.

Примеры проектов

Смоделируете систему предсказания спроса с учётом неопределённости. Разработаете вероятностную модель, оценивающую спрос на товары с учётом сезонности и случайных факторов. Проведёте анализ исторических данных, построите распределения вероятностей, примените методы Монте-Карло для прогнозирования и оцените доверительные интервалы. Визуализируете результаты и сделаете выводы о стабильности модели.

Инструменты

Python, NumPy, SciPy, Statsmodels, Matplotlib, Seaborn.

Специальные разделы математической статистики

МИФИ

3 з. е.

Познакомьтесь с основами математической статистики и методами проверки гипотез. Научитесь работать с основными распределениями, рассчитывать доверительные интервалы и анализировать дисперсию. Освойте методы оценки параметров и проверки статистических зависимостей. Разберётесь в ключевых подходах к анализу данных, научитесь применять их в задачах машинного обучения и инженерии данных.

Примеры проектов

Проведёте статистический анализ факторов, влияющих на успеваемость студентов. Соберёте и обработаете данные о студентах, их занятиях, оценках и внешних факторах. Исследуете основные статистические закономерности, примените ANOVA, регрессионный анализ и тесты гипотез для выявления значимых факторов. Рассчитаете доверительные интервалы и проверите гипотезу о влиянии учебных привычек на итоговые результаты. Оформите выводы в виде аналитического отчёта с визуализациями.

Инструменты

Python, NumPy, SciPy, Statsmodels, Matplotlib, Seaborn.

А/В-тестирование

МИФИ

3 з. е.

Научитесь готовить и проводить А/В-тестирование. Разберётесь, как найти баланс между получением полезных результатов и затраченными ресурсами. Узнаете, как корректно использовать А/В-тестирование при принятии решений.

Примеры проектов

Проведёте А/В-тестирование нового продукта в аудиосервисе. Возьмёте на себя роль аналитика приложения для прослушивания аудиокниг. Определите целевую аудиторию для продвижения новой подписки на детские аудиокниги и спроектируете эксперимент. Рассчитаете ключевые метрики, выберете оптимальный объём выборки и оцените статистическую значимость результатов. Проведёте анализ данных, сформируете портрет пользователя и обоснуете бизнес-решение на основе А/В-теста.

Инструменты

Python, NumPy, SciPy, Statsmodels, Matplotlib, Seaborn.

2 семестр

Алгоритмы и структуры данных

Практикум

7 з. е.

Познакомитесь с принципами применения алгоритмов для решения различных задач. Научитесь находить и реализовывать альтернативные способы решения, оценивать их эффективность. Разберётесь в ключевых структурах данных и освоите их применение для оптимального решения конкретных задач.

Примеры проектов

Разработаете поисковую систему для текстовых документов. Создадите поисковый алгоритм, который индексирует текстовые файлы и позволяет быстро находить нужные документы по ключевым словам. Реализуете основные структуры данных (хэш-таблицы, деревья, графы) для хранения индексов. Оптимизируете поиск с помощью префиксных деревьев (Trie) или обратного индекса. Оцените производительность алгоритма и сравните различные подходы. Оформите код в виде модульного проекта и протестируете его.

Инструменты

Python.

Машинное обучение без учителя

МИФИ

3 з. е.

Научитесь применять машинное обучение для наборов данных, не имеющих явного признака для предсказания. Выучите практики, позволяющие кластеризовать данные, уменьшать их размерность и находить аномалии.

Примеры проектов

Разделите покупателей магазина на сегменты и поймёте, чем эти сегменты между собой отличаются.

Инструменты

Scikit-learn.

Учебная практика

МИФИ

2 з. е.

Программа дисциплины поможет углубить и применить на практике полученные теоретические знания. Вы получите опыт работы с реальными данными от заказчиков. Сможете развить навыки организации процесса и самостоятельного принятия решений. Соберёте кейсы для портфолио и защитите свой проект.

Принципы разработки на Python

МИФИ

5 з. е.

Познакомьтесь с принципами разработки промышленного программного обеспечения. Научитесь проектировать будущие программы, работать с системами контроля версий и управлять зависимостями. Освойте практики тестирования, отладки и развёртывания.

Примеры проектов

Напишите программу, позволяющую автоматизировать рутинные операции на рабочем месте.

Инструменты

Python, VS Code, flake8, Pylint.

Операционные системы семейства Unix

МИФИ

4 з. е.

Познакомьтесь с основами операционных систем семейства Unix и их архитектурой. Научитесь работать с файловой системой, процессами и управлением ресурсами. Освойте базовые команды Shell, автоматизацию задач с помощью Bash-скриптов, а также разберётесь в механизмах многозадачности, сетевого взаимодействия и безопасности. Научитесь настраивать окружение, управлять пользователями и анализировать работу системы.

Примеры проектов

Настройте и автоматизируете работу серверного окружения. Развернёте виртуальную машину с Linux, настройте пользователей, права доступа и сетевое взаимодействие. Создадите Bash-скрипт для автоматизации установки программ, мониторинга системы и логирования ошибок. Настройте SSH-доступ и разработаете план резервного копирования.

Инструменты

Bash, Shell scripting, SSH.

Прогнозирование временных рядов

Практикум

5 з. е.

Временные ряды описывают, как параметры, например, объём потребления электроэнергии или количество заказов такси, меняются с течением времени. Вы научитесь анализировать ряды, искать тренды и выявлять сезонность. Узнаете, как создавать табличные данные и применять регрессионный анализ для исследования временных рядов.

Примеры проектов

Построите модель и спрогнозируете пиковые нагрузки в такси.

Инструменты

Анализ временных рядов, задача регрессии.

Навыки коммуникации и представления данных

Практикум

3 з. е.

Научитесь ясно и убедительно доносить техническую информацию до разных аудиторий. Освойте сторителлинг в Data Science и эффективную визуализацию данных. Научитесь выстраивать структуру презентации, подбирать подходящие графики и уверенно отвечать на вопросы. Разовьёте навыки командной коммуникации и объяснения ML-моделей как техническим, так и бизнес-специалистам.

Примеры проектов

Разработаете решение для продуктовой команды банка, которое поможет сократить отток новых клиентов после регистрации в мобильном приложении. Вы сформулируете бизнес-задачу, трансформируете её в ML-формат, определите метрики эффективности, и адаптируете сложные технические детали под бизнес-аудиторию.

Инструменты

Деловая коммуникация, JTBD, SWOT, Pain-Gain Map, управление требованиями, адаптация технической информации, навыки переговоров, подготовка презентаций.

3 семестр

Работа с изображениями и текстами

Практикум

3 з. е.

Поймёте принципы работы с текстовыми и медиаданными (изображения, видео). Разбёрете, чем их предобработка отличается от табличных данных. Познакомитесь с задачами, которые решаются в областях NLP и CV, и изучите базовые подходы к их решению.

Примеры проектов

Научитесь классифицировать изображения и тексты для разных бизнес-задач.

Инструменты

OpenCV, Pillow, PyTorch, NLTK, spaCy.

Классические и глубинные модели по работе с текстом

МИФИ

3 з. е.

Познакомитесь с классическими и глубинными методами обработки текста. Научитесь работать с моделями на основе статистических подходов, векторных представлений слов и нейросетевых архитектур. Освоите различные рекуррентные сети. Проведёте анализ текстов и их классификацию. Разберётесь в принципах работы современных NLP-моделей и их применении в реальных задачах.

Примеры проектов

Разработаете систему автоматической классификации отзывов для онлайн-магазина. Разработаете и обучите NLP-модель для автоматической классификации отзывов клиентов.

Инструменты

spaCy, TF-IDF, RNN, LSTM, BeautifulSoup.

Нейросетевые модели для последовательностей

МИФИ

3 з. е.

Познакомитесь с основными подходами к моделированию последовательностей в NLP. Научитесь работать архитектурами моделей для машинного перевода, текстовой суммаризации и генерации. Разберётесь в механизме внимания и обучите модели для обработки последовательностей.

Примеры проектов

Разработаете модель для автоматической суммаризации новостных статей.

Инструменты

PyTorch, Attention, seq2seq.

Модели на базе архитектуры трансформер

МИФИ

4 з. е.

Познакомитесь с ключевыми принципами архитектуры трансформеров и их применением в задачах обработки естественного языка. Научитесь работать с предобученными моделями, такими как BERT, адаптировать их под конкретные задачи с помощью дообучения.

Примеры проектов

Используете предобученную модель BERT, адаптируете её под ваш корпус данных.

Инструменты

Hugging Face, Transformers, PyTorch.

Большие языковые модели

МИФИ

3 з. е.

Познакомитесь с принципами работы больших языковых моделей (LLM) и их архитектурами, такими как GPT. Научитесь дообучать и адаптировать предобученные модели под специфические задачи.

Примеры проектов

Обучите и доработаете предобученную LLM-модель на юридических текстах, реализуете механизм обработки длинных контекстов.

Инструменты

LLM, Fine-Tuning.

Генеративные модели для речи и текста

МИФИ

3 з. е.

Познакомитесь с принципами работы генеративных моделей в обработке естественного языка. Научитесь обучать и дообучать трансформеры, создавать текстовые модели для генерации, суммаризации и дополнения текста. Освоите методы оценки качества генерации.

Примеры проектов

Разработаете текстовую модель для автоматического написания кратких аннотаций к новостям.

Инструменты

Hugging Face, PyTorch, BLEU.

Учебная практика

МИФИ

7 з. е.

Программа дисциплины поможет углубить и применить на практике полученные теоретические знания. Вы получите опыт работы с реальными данными от заказчиков. Сможете развить навыки организации процесса и самостоятельного принятия решений. Соберёте кейсы для портфолио и защитите свой проект.

Гибкие методологии

Практикум

3 з. е.

Познакомитесь с гибкими методологиями разработки, научитесь организовывать процессы в команде и адаптироваться к изменениям. Освоите бизнес-коммуникации, научитесь понимать заказчика и формулировать задачи. Разовьёте критическое мышление, научитесь анализировать информацию, аргументировать решения и правильно эскалировать проблемы. Освоите навыки презентации, научитесь понятно объяснять сложные идеи и работать с визуализацией данных. Будете эффективно взаимодействовать с командой и бизнесом, разберётесь в основах управления людьми и проектами.

Примеры проектов

Проанализируете существующие рабочие процессы команды, выявите узкие места и предложите способы оптимизации. Настроите систему управления задачами с учётом гибких методологий. Проведёте интервью с «заказчиком» для уточнения требований. Разработаете план улучшения коммуникации в команде и представите его в формате визуализированной презентации. Научитесь аргументировать решения перед бизнесом и анализировать их эффективность.

Инструменты

Agile, Scrum, Kanban, управление процессами, выявление требований, переговоры, стресс-менеджмент, критическое мышление, презентация данных, сторителлинг, управление командой, лидерство.

4 семестр

Производственная практика (научно-исследовательская работа)

Практикум и МИФИ 22 з. е.

Преддипломная практика

Практикум и МИФИ 3 з. е.

Выполнение и защита выпускной квалификационной работы

МИФИ 6 з. е.