

Специалист по работе с данными и применению ИИ Data-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 Engineering

29 з. е.

- Алгоритмы и структуры данных 7 з. е.
- Анализ бизнес-метрик и продвинутый SQL 5 з. е.
- Практическая бизнес-аналитика 3 з. е.
- Введение в системный анализ 5 з. е.
- Операционные системы семейства Unix 4 з. е.
- Учебная практика 2 з. е.
- Навыки коммуникации и представления данных 3 з. е.

3 семестр

Углублённый Data Engineering

29 з. е.

- Работа с данными в хранилище 3 з. е.
- Автоматизация подготовки данных 3 з. е.
- Принципы проверки качества входных данных 3 з. е.
- Аналитические базы данных и хранилища 4 з. е.
- Поточковая обработка данных и их хранение в озере 3 з. е.
- Облачные технологии и проектный практикум 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.

Анализ бизнес-метрик и продвинутый SQL

Практикум

5 з. е.

Познакомитесь с ключевыми бизнес-метриками и научитесь анализировать их с помощью SQL. Освоите методы расчёта конверсий, когортного и ретеншн-анализа, юнит-экономики. Разберётесь в сложных SQL-запросах, включая оконные функции, CTE, подзапросы и агрегированные вычисления.

Примеры проектов

Создадите набор SQL-запросов для расчёта бизнес-метрик: конверсии, среднего чека, LTV и ретеншн-анализа.

Инструменты

Метрики (CTR, Retention и другие) SQL.

Практическая бизнес-аналитика

Практикум

3 з. е.

Познакомитесь с основами бизнес-анализа и подходами к решению бизнес-задач. Научитесь выявлять потребности заинтересованных сторон, формулировать требования и анализировать бизнес-процессы. Освоите методы сбора и обработки данных, построения моделей процессов и оценки эффективности решений. Разберётесь в инструментах бизнес-аналитики и научитесь применять их для оптимизации бизнеса и принятия стратегических решений.

Примеры проектов

Поработаете с интервью с заказчиком по вопросам внутренней логистики. Проанализируете ASIS, разработаете TOBE. Проанализируете и разработаете автоматизацию процессов.

Инструменты

Draw.io, UML, BPMN, IDEF0.

Введение в системный анализ

Практикум

5 з. е.

Познакомитесь с принципами системного подхода и научитесь анализировать сложные системы с точки зрения их структуры, функций и взаимодействия компонентов. Освоите методы моделирования систем, выявления их слабых мест и оптимизации процессов. Узнаете, как проектировать архитектуру систем, учитывая технические и функциональные требования. Разберётесь в подходах к интеграции систем, управлению изменениями и обеспечению их устойчивости. Научитесь применять системное мышление для решения бизнес-задач.

Примеры проектов

Разработаете архитектуру системы обработки данных для логистической компании. Проанализируете текущие бизнес-процессы компании и выявите узкие места. Смоделируете бизнес-процесс с помощью BPMN и сформируете требования к системе обработки данных. Создадите UML-диаграммы для описания структуры системы. Опишете варианты интеграции с внешними сервисами и разработаете концептуальную архитектуру будущей системы.

Инструменты

Draw.io, UML, BPMN, IDEF0.

Операционные системы семейства Unix

МИФИ

4 з. е.

Познакомитесь с основами операционных систем семейства Unix и их архитектурой. Научитесь работать с файловой системой, процессами и управлением ресурсами. Освоите 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

Построите модель и спрогнозируете пиковые нагрузки в такси.

Инструменты

Bash, Shell scripting, SSH.

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.

Навыки коммуникации и представления данных

Практикум

3 з. е.

Научитесь ясно и убедительно доносить техническую информацию до разных аудиторий. Освойте эффективную визуализацию данных и сторителлинг. Научитесь выстраивать структуру презентации, подбирать подходящие графики и уверенно отвечать на вопросы. Разовьёте навыки командной коммуникации и объяснения задач и решений как техническим специалистам, так и менеджерам с заказчиками.

Примеры проектов

Разработаете решение для продуктовой команды банка, которое поможет сократить отток новых клиентов после регистрации в мобильном приложении. Вы сформулируете бизнес-задачу, 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 данными в хранилище

Практикум

3 з. е.

Познакомитесь с основными технологиями, необходимыми для строительства хранилищ данных. Изучите различные подходы к их проектированию и научитесь работать с требованиями заказчика, выбирая оптимальные решения для конкретных задач. Разберётесь в концепции витрин данных, освоите их построение и обновление. Овладеете методами инкрементальной загрузки и транзакционной обработки. Узнаете, как оптимизировать запросы для повышения эффективности работы с данными.

Примеры проектов

Построите витрину с инкрементальной загрузкой для аналитики аудитории интернет-магазина. Приведёте в порядок модель данных и осуществите миграцию в текущем хранилище.

Инструменты

SQL, PostgreSQL.

Автоматизация подготовки данных

Практикум

3 з. е.

Автоматизируете пайплайн работы с данными: настроите автоматическую выгрузку данных из источников и освоите методы их регулярной и инкрементальной загрузки в базу данных.

Примеры проектов

Построите для e-commerce проекта пайплайн автоматизированного получения, обработки и загрузки данных (ETL) от источников до витрины.

Инструменты

Python, Airflow, PostgreSQL.

Принципы проверки качества входных данных

Практикум

3 з. е.

Разберёте принципы оценки качества данных и настроите шаги, необходимые для проверки.

Примеры проектов

Разработаете модель для автоматической суммаризации новостных статей.

Инструменты

PyTorch, Attention, seq2seq.

Аналитические базы данных и хранилища

Практикум

4 з. е.

Построите DWH с нуля на реляционной СУБД и познакомитесь с MongoDB как источником данных. Изучите организацию хранилища в Vertica, освоите базовые операции с данными и создадите простое хранилище данных в Vertica.

Примеры проектов

Спроектируете и реализуете DWH для стартапа. Построите DWH для высоконагруженной системы данных с использованием Vertica.

Инструменты

PostgreSQL, MongoDB, Vertica, Airflow.

Потоковая обработка данных и их хранение в озере

Практикум

3 з. е.

Познакомитесь с архитектурой Data Lake и принципами его наполнения данными из различных источников. Освоите обработку данных в MPP-системах и научитесь использовать PySpark и Airflow для управления обработкой данных. Разберётесь в особенностях потоковой обработки и построите свою стриминговую систему. В завершение курса создадите витрину данных с использованием real-time потоков.

Примеры проектов

Построите Data Lake для сбора данных о просмотрах видео. Автоматизируете загрузку и обработку данных в нём. Реализуете real-time обработку данных с помощью Kafka и Spark Streaming, чтобы анализировать активность пользователей, выявлять тренды и аномалии.

Инструменты

Hadoop, MapReduce, HDFS, Apache Spark (PySpark), Kafka, Spark Streaming.

Облачные технологии и проектный практикум

Практикум

3 з. е.

Познакомитесь с принципами облачных технологий и разберётесь в инфраструктуре Yandex Cloud. Научитесь развёртывать DWH, работать с Redis и Kafka для хранения и потоковой обработки данных. Освоите контейнеризацию в Docker, управление микросервисами через Kubernetes и настройку сервисов с Helm. Напишете свой первый облачный сервис с PostgreSQL, подготовите окружение и логику работы. Разберётесь в визуализации данных с DataLens и научитесь создавать интерактивные дашборды. В завершение курса подготовите инфраструктуру для хранения и обработки данных в облаке.

Примеры проектов

Разработаете инфраструктуры хранения и обработки данных в облаке.

Инструменты

Яндекс Облако, Kubernetes, Kubectl, Redis, PostgreSQL.

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.