Программа курса по Data-science:

1 месяц - Введение в основы языка и математики:

- Основы синтаксиса и базовые конструкции Python: Вы освоите условия в языке Python, структуры данных в Python (Списки, Кортежи, Множества и Словари) и функции, а также научитесь работать с файлами.

- Линейная Алгебра: Линейная алгебра - это раздел математики, который чрезвычайно полезен в Data Science и машинном обучении. Она используется при предварительной обработке данных, в преобразовании данных и оценке моделей. На уроках вы изучите векторы, матрицы и решения систем линейных уравнений.

2 месяц - Введение в основные инструменты Data Science:

- Основы мат-анализа: Дифференциальное и интегральное исчисление лежат в основе Data Science и машинного обучения. В анализе данных он используется в основном для оптимизации — подбора наилучших параметров системы для минимизации или максимизации целевой функции. Вы освоите производные и интегралы и научитесь применять их на практике.

- Работа с данными: Продвинутые операции и работа с отсутствующими данными.

- Введение в NumPy: NumPy - это библиотека для языка программирования Python, которая используется для работы с числами и матрицами (двумерными таблицами). Она делает вычисления с числами намного быстрее и удобнее, особенно если нужно работать с большими объемами данных. На уроках вы освоите базовые операции, продвинутые операции и работу с матрицами в NumPy.

- Введение в Pandas: Pandas - это это библиотека для Python, которая помогает работать с таблицами и упрощает обработку данных. В Pandas можно удобно хранить и манипулировать табличными данными, как в Excel, только это быстрее и лучше для программирования. Вы изучите базовые операции, продвинутые операции и работу с отсутствующими данными.

- Введение в Matplotlib: Matplotlib - это библиотека для Python, которая используется для создания графиков и визуализации данных. Она помогает представлять данные в виде графиков, диаграмм и других изображений, чтобы лучше понимать информацию. Вы научитесь построению графиков с помощью библиотеки Matplotlib.

- Введение в Seaborn: Создание графиков, основные понятия статистики, распределения вероятностей, тестирование гипотез, корреляция и регрессия.

- Введение в базы данных и SQL: SQL – это язык запросов, который используется для работы с базами данных.

- Соединения таблиц и группировка: Работа с табличными данными, их оптимизация и группировка по необходимым категориям.

3 месяц - Введение в статистику и машинное обучение:

- Введение в Power BI: Импорт данных, Очистка и преобразование данных с Power Query. Создание визуализаций, фильтры, срезы и отчеты, Основы DAX для расчетов и анализа. Публикация и совместная работа в Power BI Service.

- Основные понятия статистики: Изучение статистики и ее понятий, распределения вероятностей, тестирование гипотез, корреляция и регрессия.

4 месяц - Введение в обучение и нейронные сети:

- Введение в машинное обучение: Введение в подраздел искусственного интеллекта, который использует техники для обучения компьютерных систем выполнять задачи.

- Алгоритмы решений: Вы освоите деревья решений и алгоритм случайного леса.

- Метод опорных векторов (SVM): SVM - один из наиболее популярных методов обучения, который применяется для решения задач классификации и регрессии. С помощью этого метода можно классифицировать текст или изображения, обнаружить спам, распознавать лица, делать прогнозы и так далее.

- Кластеризация: Кластеризация - это задача группировки множества объектов на подмножества с использованием алгоритма k-means. Она помогает упорядочить данные по различным группам в зависимости от особенностей.

- PCA и снижение размерности: PCA (Principal Component Analysis) - это один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации. Меньшие наборы данных легче исследовать и визуализировать, а анализ данных становится намного проще и быстрее для алгоритмов машинного обучения без обработки посторонних переменных.

- Оценка моделей и Cross-Validation: Cross-Validation - это методика обучения и оценки модели, которая разбивает данные на несколько секций и обучает несколько алгоритмов на этих секциях. Вы научитесь анализу главных компонентов для упрощения данных и оценке модели с помощью Cross-Validation.

5 месяц - Основы Big Data:

- Введение в глубокое обучение: Глубокое обучение рекуррентные и сверточные нейронные сети, применение их в различных задачах.

- RNN и CNN: Изучение рекуррентных нейронных сетей(RNN), Сверточных нейронных сетей (CNN).

- Введение NLP: NLP - Одно из направлений искусственного интеллекта, которое работает с анализом, пониманием и генерацией живых языков. Благодаря NLP компьютеры могут читать, интерпретировать, понимать человеческий язык, а также выдавать ответные результаты.

- Введение в Big data: Big Data - это гигантские массивы информации, которую генерируют большие информационные системы. Вы изучите модели глубокого обучения для больших данных.

- Практический урок по Data Science: Анализ данных о пассажирах Titanic; Введение в обработку естественного языка.

- Финальный проект (Часть 1): Завершение проекта. Подведение итогов и обсуждение карьерных возможностей в Data Science.

6 месяц - Продвинутые темы Data Science:

- Анализ временных рядов: Статистический метод анализа данных, состоящих из последовательности значений, собранных через равные промежутки времени.

- Практический урок по Data Science: Введение в Big Data и распределенные вычисления.

- Рекомендательные системы: Рекомендательная система — это тип системы машинного обучения, которая предоставляет пользователям персонализированные рекомендации на основе их прошлого поведения, предпочтений и шаблонов. Вы научитесь внедрять рекомендательные системы, используя контентные и коллаборативные методы фильтрации.

- XAI: XAI - это модель, которая могла бы в перспективе объяснять механизмы, лежащие за алгоритмами машинного обучения.

- AutoML: AutoML - это автоматизированное машинное обучение. Инструменты AutoML упрощают процесс обработки данных, используя имеющуюся информацию. Вы научитесь автоматизации построения и оптимизации моделей.

- Развертывание моделей: Развертывание модели — это процесс внедрения моделей машинного обучения в производство. Вас ждет практика по развертыванию модели в реальных условиях с помощью Flask, Docker и REST API, а также работа с Apache Spark для обработки больших данных.

- Обработка больших данных с Apache Spark: Apache Spark — это фреймворк для обработки и анализа больших объёмов информации, входящий в инфраструктуру Hadoop.

- Подведение итогов: Обсуждение карьерных возможностей в Data Science, Этика и ответственность в Data Science.