

Учебный план дополнительной профессиональной программы переподготовки:

«Наука о данных: комплексная подготовка от базовых алгоритмов до интеграции нейросетей в ИТ-ландшафт»

№ п/п	Наименование темы	Описание темы	Всего, час	В том числе			
				Лекции	Практ. раб.	Сам. работа	Аттест.
Модуль 1 «Основы программирования и баз данных»							
1.1	Базовые понятия и определения	Прикладная задача – постановка и решение. Алгоритм – свойства и способы представления. Программа – подготовка и исполнение на компьютере. Данные – способы представления и обработки. Типы данных – назначение и роль в программе. Операнды и операторы – вычисление выражений. Основы программирования на Python. Интерактивный и пакетный режимы работы Python. Переменные и константы. Типы данных – изменяемые и неизменяемые. Работа с целыми числами и числами с плавающей точкой. Базовые операции работы со строками. Простые операции ввода-вывода. Преобразование типов. Разработка программ линейной структуры: ввод – обработка – вывод.	12	6	2	4	-
1.2	Разработка программ: модели и методологии	Методологии и языки программирования – обзор. Императивное и декларативное программирование. Структурное и объектно-ориентированное. Функциональное и	10	4	4	2	-



		логическое. Структурное программирование. Базовые принципы: блочная структура кода – блоки и подпрограммы, типовые структуры управления – последовательность, ветвление, цикл, «нисходящая» разработка – пошаговая детализация программы. Структурное программирование на Python. Инструкции ветвления и цикла. Инструкции break, continue и pass. Функции. Структура программы. Наборы данных – кортежи, списки, словари, множества. Базовые принципы: абстрагирование, инкапсуляция, наследование, полиморфизм. Объектно-ориентированное программирование на Python. Объектная модель Python – встроенные типы как объекты. Объявление классов и создание объектов. Наследование классов. Полиморфизм и отделение интерфейса от реализации. Концепция MVC. Программирование типовых алгоритмов обработки наборов данных. Объектно-ориентированное программирование					
1.3	Работа с данными из внешних источников	Введение в Python, среды исполнения (IDE). Типы данных в Python. Базовые конструкции и структуры. Циклы и условия. Функции и классы. Библиотеки Numpy, Scipy для научных вычислений. Библиотека Pandas, как стандарт исследования данных. Визуализация в	8	4	4	-	-

		Python и срезы данных. Библиотеки Matplotlib, seaborn, plotly. Работа с командной строкой. Linux, Windows. Решение практического кейса: «Визуализация данных с использованием основных библиотек языка программирования Python в среде Jupyter Notebook».					
1.4	Промежуточная аттестация	Тестирование	2	-	-	2	-
Модуль 2 «Наука о данных. Уровень 1: основы работы с данными (Data Science)»							
2.1	Основы программирования на языке Python для науки о данных	Основы языка: переменные и их типы, арифметические операции, операции над строками, функции. Основы объектно-ориентированного программирования: классы, экземпляр класса, поля и методы, наследование, безопасный доступ. Библиотеки Numpy, Scipy: векторы, матрицы, операции над векторами и матрицами, основы линейной алгебры. Работа в среде PyCharm – создание законченных скриптов. Работа в среде Google Colabs – Python notebooks.	18	6	10	2	-
2.2	Основы работы с базами данных	Основные типы баз данных в науке о данных: основы хранения информации, таблицы и коллекции, отношения между ними, транзакции, уровни изоляции. Реляционные СУБД (РСУБД) как основной источник данных в науке о данных: разбор основных возможностей РСУБД на примере СУБД	16	4	10	2	-

		«PostgreSQL». NoSQL (Not Only SQL – не только SQL) базы данных на примере документоориентированной системы управления базами данных MongoDB: коллекции, операции по созданию, изменению, вставке и удалению данных. Базы данных, хранящиеся в ОЗУ на примере базы «Redis»: хранение информации «ключ-значение», операции вставки, манипуляции и удаление. Примеры работы с основными типами БД из языка программирования Python из баз данных «PyCharm» и «Python Notebook».					
2.3	Основы анализа данных	Основы работы с библиотеками Request, BeautifulSoup — сбор данных с HTML-страниц. Библиотеки Pandas и Polars: универсальные табличные процессоры без ограничений. Библиотеки Matplotlib и Seaborn: основы визуализации данных. Exploratory Data Analysis (EDA) — базовые принципы анализа данных с использованием графиков. Основы бизнес-анализа и метода анализа прибыльности бизнеса, основанного на расчете доходов и расходов на одну единицу продукции: анализ данных и расчёт бизнес-метрик: пожизненная ценность клиента, метод сегментации клиентов, средний доход, получаемый компанией от одного активного пользователя за определенный период,	22	6	12	4	-

		средний доход, полученный от одного платящего пользователя за определенный период и т. д.					
2.4	Математика науки о данных	Элементарные функции: основные определения, предел функции. Непрерывность и дифференцируемость: производные простых и сложных функций, оптимизация функции одной переменной. Теория вероятностей: сложение и произведение вероятностей, формула Бернулли, условная вероятность: теорема Байеса. Математическая статистика: проверка гипотез, статистический вывод, центральная предельная теорема, сравнение средних значений, корреляция, регрессия, открытые коды, АБ-тесты.	14	2	10	2	-
2.10	Промежуточная аттестация	Тестирование	2	-	-	2	-
Модуль 3 «Наука о данных: профессиональный уровень (Data Science PRO)»							
Тема 1. Базовый модуль. Большие данные							
3.1.1	Введение в большие данные. Знакомство с основными понятиями. Статистика	Введение в предмет. Какие бывают источники данных, характеристики, корреляция. Типы структурированности данных. Введение в статистику. Типы распределений данных. Проверка гипотез, матрицы ошибок Структуры данных. Техническое задание и требование к проекту больших данных, науки о данных.	11	8	-	3	-

3.1.2	Базы данных, их типы и моделирование. Взаимодействие пользователей с данными	Системы хранения данных. Базы данных. Реляционные базы данных и их моделирование. Программа PGAdmin. Язык программирования SQL, синтаксис обращения к базам данных. Простые и сложные запросы. Решение практического кейса: «Создание базы данных университета».	16	3	10	3	-
3.1.3	Языки программирования, среды исполнения и основные операционные системы, которые используются в задачах больших данных	Введение в Python, среды исполнения (IDE). Типы данных в языке программирования Python. Базовые конструкции и структуры. Циклы и условия. Функции и классы. Библиотеки Numpy, Scipy для научных вычислений. Библиотека Pandas, как стандарт исследования данных. Визуализация в Python и срезы данных. Библиотеки Matplotlib, seaborn, plotly. Работа с командной строкой. Операционные системы Linux, Windows. Решение практического кейса: «Визуализация данных с использованием основных библиотек языка программирования Python в среде «Jupyter Notebook».	24	4	11	9	-
3.1.4	Основная последовательность машинного обучения и потоков данных. Облачные платформы	Последовательность машинного обучения. Особенности архитектурных решений. Sklearn - основная библиотека машинного обучения. Ленивые вычисления. Решение практического кейса: «Создание последовательности действий для проекта по машинному обучению с использованием библиотеки Sklearn».	12	3	4	5	-



3.1.5	Автоматизированный сбор и структурирование данных из открытых источников с помощью скриптов, подготовка выборки, преобразование и балансировка данных	Ключевые задачи в подготовке датасетов и их важность. Нормализация и стандартизация. Поиск аномалий и выбросов. Методы обработки и визуализации. Разбалансированные датасеты и методы балансировки. Доверительные интервалы. Обработка категориальных признаков. Инструменты LabelEncoder, One Hot encoding. Решение практического кейса: «Преобразование данных для подготовки датасета к решению задач машинного обучения».	16	5	5	6	-
3.1.6	Обучение без учителя. Работа с размерностью и структурой данных	Обучение на размеченных и размеченных данных. Проблема снижения размерности данных. Метод главных компонент. Алгоритмы кластеризации. Метод k-средних, c-средних, иерархическая кластеризация. Выбор алгоритма кластеризации. Решение практического кейса: «Снижение размерности в задаче определения кредитного потенциала заемщика» Решение практического кейса: «Кластеризация городов Российской Федерации по типу прироста (убыли) населения».	14	2	7	5	-
3.1.7	Теория вероятности и вероятностные классификаторы	Полная и условная вероятность, теорема Байеса Байесовский вероятностный классификатор. Решение практического кейса: «Решение задачи классификации осколков стекла по их химическому составу».	7	3	2	2	

3.1.8	Классификация. Оценка качества алгоритма	Задача классификации. Матрица ошибок (Confusion -matrix). Точность классификации. Модель анализа качества для задач бинарной классификации, показывающая способность модели разделять классы. Методы классификации. Метод ближайших соседей. Метод опорных векторов. Решение практического кейса: «Оценка и классификация кредитного потенциала заемщиков».	12	4	5	3	
3.1.9	Ансамбли и повышение точности алгоритмов	Ансамбли решателей. Решающие деревья. Случайный лес. Ансамбли алгоритмов. Алгоритмы улучшения стабильности и точности алгоритмов. Основы составления сеток обучения. Решение практического кейса: «Решение задач медицинской диагностики с применением ансамблей классификаторов» Решение практического кейса: «Решение задачи анализа сетевой активности».	11	3	6	2	
3.1.10	Регрессия	Линейная регрессия. Логистическая регрессия. Метод наименьших квадратов. Средняя квадратичная ошибка, средняя абсолютная ошибка. Авторегрессия. Решение практического кейса: «Прогнозирование объемов продаж товаров торгового предприятия». Решение практического кейса: «Оценка кредитоспособности клиентов банка - физических лиц».	23	3	12	8	

3.1.11	Нейронные сети	Введение в искусственные нейронные сети. Биологическая модель нейрона. Активационные функции. Полносвязные нейронные сети. Методы обучения нейронных сетей. Борьба с переобучением. Нейронные сети Кохонена. Сверточные нейронные сети. Решение практического кейса: «Моделирование объемов продаж торгового предприятия с использованием нейронных сетей». Решение практического кейса: «Анализ банковского сектора с применением нейронных сетей Кохонена».	16	5	7	4	
3.1.12	Промежуточная аттестация	Тестирование	10	-	-	10	-

Тема 2. Профильный модуль. Нейронные сети

3.2.1	Разработка компьютерных программ: написание программного кода и решение практических задач с применением нейронных сетей	Решение практического кейса: «Разработка приложения на языке программирования Python для решения задачи компьютерного зрения с использованием нейронных сетей».	25	0	16	9	-
3.2.2	Рекомендательные системы	Введение. Коллаборативная фильтрация и фильтрация на основе содержания. Сингулярное разложение матрицы. Достоинства и недостатки. Решение практического кейса: «Создание рекомендательной системы торгового предприятия».	11	2	6	3	-
3.2.3	Создание приложений, интеграция в ИТ	Сортирование и сохранение моделей. Работа с предобученными	16	3	8	5	-

	ландшафт	моделями. Фласк-приложение (Flask-приложение). Выведение моделей в производство. Контроль версии моделей с серверной архитектурой. Облачная платформа Kaggle для решения задач науки о данных. Решение практического кейса: «Создание фласк-приложения (Flask-приложения) для запуска моделей машинного обучения».					
3.2.4	Промежуточная аттестация	Тестирование	10	-	-	10	-
4	Итоговая аттестация	Защита ВКР	28	-	-	-	28
	ИТОГО		366	80	151	107	28