**Глава I**

**Рабочая программа**

Курс носит междисциплинарный характер и может быть фактически разнесен между часами, отведенными на элективные дисциплины и внеурочную деятельность.

Предлагаемая программа соответствует положениям федерального государственного образовательного стандарта среднего общего образования.

Рабочая программа составлена на основе:

* Закона об образовании Российской Федерации
* Федерального государственного образовательного стандарта среднего общего образования.
* Профессионального стандарта «06.001 Разработка программного обеспечения»
* Профессионального стандарта «08.022 Статистическая деятельность»

**ЦЕЛИ И ЗАДАЧИ РЕАЛИЗАЦИИ ОСНОВНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ ОСНОВНОГО ОБЩЕГО ОБРАЗОВАНИЯ ПО КУРСУ**

Глобальными целями курса являются формирование у обучающегося аналитического мышления и, соответственно, знаний и умений, необходимых для успешного развития в отраслях, связанных со сложной аналитикой данных. Для достижение поставленных целей образование по данному направлению должно обеспечить решение следующих задач:

* овладение реальными и практическими знаниями методов статистического анализа данных;
* формирование навыков построения математических моделей (от нейронных сетей до кластеризации, от факторного до корреляционного анализа);
* формирование навыков работы с большими массивами данных;
* осознание практической важности нахождения уникальной закономерности в данных.

**ПРИНЦИПЫ И ПОДХОДЫ К ФОРМИРОВАНИЮ ПРОГРАММЫ**

Стандарт второго поколения (ФГОС) в сравнении со стандартом первого поколения предполагает деятельностный подход к обучению, где главная цель: развитие личности учащегося. Система образования отказывается от традиционного представления результатов обучения в виде знаний, умений и навыков. Формулировки стандарта указывают реальные виды деятельности, которыми следует овладеть к концу обучения, т. е. обучающиеся должны уметь учиться, самостоятельно добывать знания, анализировать, отбирать нужную информацию, уметь контактировать в различных по возрастному составу группах. Data scientist – это профессия, которая требует реальных и практических знаний методов статистического анализа данных, навыков построения математических моделей (от нейронных сетей до кластеризации, от факторного до корреляционного анализа), работы с большими массивами данных и уникальной способности находить закономерности. Будущему специалисту важно обладать не только фундаментальными, инженерными знаниями, но и разбираться в предметной области, в которой будет идти работа.

**СОСТАВ УЧАСТНИКОВ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА**

Программа основного общего образования рассчитана на реализацию в 10 - 11 классах общеобразовательных учреждений и учреждений с углубленным изучением отдельных предметов, и нацелена на возрастную категорию учащихся 15 - 18 лет.

**ОБЩАЯ ХАРАКТЕРИСТИКА УЧЕБНОГО КУРСА**

Представленная программа направления «Большие данные (10-11 класс)» предназначена для практического освоения учащимися работы с технологиями информационного поиска и обработки больших данных, работы с инструментами анализа данных, основ математической статистики и теории вероятностей, основ математического моделирования. Программа рассчитана на 2 года (10-11 класс), при этом обучение можно разделить на 4 модуля:

* Модуль «Введение в вероятностное моделирование» содержит основы исчисления вероятностей, вероятностного анализа данных и начальные сведения о вероятностных моделях, использующихся для решения задач машинного обучения. В рамках модуля излагаются примеры применения изучаемых моделей, методов и алгоритмов, а также типовые алгоритмы решения задач реального мира с использованием вероятностных методов и моделей.
* Модуль «Анализ и визуализация данных на Python» предполагает изучение основных методов, подходов и инструментов для анализа и визуализации данных с использованием возможностей Python и его основных библиотек.
* Модуль «Параллельная обработка и управление большими данными» предполагает изучение теории баз данных, а также современных инструментов и технологий для решения задач, связанных с параллельной обработкой и анализом больших данных.
* Модуль «Машинное обучение» предполагает изучение основных методов и моделей машинного, а также их реализацию на Python. В рамках модуля даются алгоритмы решения типовых проблем машинного обучения с примерами вариантов их применения в реальных задачах.

Обучение в классе с профилем «Data Science» даст возможность освоения некоторых базовых компетенций в этой области, а именно:

* элементы математической статистики и теории вероятностей;
* элементы математического моделирования;
* основные алгоритмы машинного обучения;
* основы программирования на языке Python;
* визуализация данных;
* основы глубокого обучения и теории искусственных нейронных сетей.

**СОДЕРЖАНИЕ КУРСА**

|  |  |  |  |
| --- | --- | --- | --- |
| № п/п | Модуль | Наименование раздела | Количество часов |
| 1 полугодие 10 класса |
|  | Введение в вероятностное моделирование | Вводное занятие. Что такое математическая модель? | 2 |
|  | Интуитивные понятия теории вероятностей. | 2 |
|  | Исчисление вероятностей и элементы комбинаторики. Текущий контроль. | 2 |
|  | Условная и полная вероятность. | 2 |
|  | Понятие случайной величины. | 2 |
|  | Обработка результатов наблюдений. Понятие статистической оценки. Текущий контроль. | 2 |
|  | Числовые оценки выборочных характеристик. | 2 |
|  | Вероятностные модели случайной величины. | 2 |
|  | Оценка параметров распределения случайной величины. Текущий контроль. | 2 |
|  | Интервальные оценки и проверка статистических гипотез. | 2 |
|  | Базовые понятия из линейной алгебры. | 2 |
|  | Элементы многомерного статистического анализа и моделирования. Базовые элементы корреляционного анализа и регрессионного анализа. Текущий контроль. | 2 |
|  | Понятие классификации и кластеризации. Как связаны эти две задачи? Чем классификация отличается от регрессии? | 2 |
|  | Понятие градиента. Текущий контроль. | 2 |
|  | Реализация итогового проекта. | 2 |
|  | Презентация результатов итогового проекта. | 2 |
| 2 полугодие 10 класса |
|  | Анализ и визуализация данных на Python | Анализ данных. Примеры и задачи. | 2 |
|  | Одномерный анализ данных. График функции. Гистограммы. Распределения. | 2 |
|  | Вектора и матрицы. Текущий контроль. | 2 |
|  | Введение в Python. Базовые операции. | 2 |
|  | Библиотека numpy. Примеры. | 2 |
|  | Библиотека pandas. Примеры. Текущий контроль. | 2 |
|  | Библиотека matplotlib. Примеры. | 2 |
|  | Понятие корреляции. Примеры на pandas и numpy. | 2 |
|  | Обучение с учителем. Примеры. Текущий контроль. | 2 |
|  | Обучение без учителя. Примеры. | 2 |
|  | Кластеризация данных на Python. | 2 |
|  | Линейная регрессия на Python. Текущий контроль. | 2 |
|  | Логистическая регрессия на Python. | 2 |
|  | Работа с изображениями в Python. Текущий контроль. | 2 |
|  | Реализация итогового проекта. | 2 |
|  | Презентация результатов итогового проекта. | 2 |
| 1 полугодие 11 класса |
|  | Параллельная обработка и управление большими данными | Понятие обработки данных. Виды обработки данных. Виды баз данных. | 2 |
|  | Типы данных, таблицы и отношения между ними. Реляционная модель данных. | 2 |
|  | Введение в SQL. Примеры в PostgreSQL. Текущий контроль. | 2 |
|  | Понятие индекса. Виды индексов. | 2 |
|  | Проектирование баз данных. Цели проектирования. Нормализация данных. Проектирование базы данных в PostgreSQL. | 2 |
|  | Текстовые и бинарные форматы хранения данных json, csv, parquet. Текущий контроль. | 2 |
|  | Обработка данных в памяти. Продвинутый pandas. Знакомство с dataframe’ами. Примеры. | 2 |
|  | Колоночные базы данных (NoSQL для больших данных): HBase, ClickHouse. | 2 |
|  | Основные понятия распределенной обработки данных. Текущий контроль. | 2 |
|  | Знакомство с Apache Spark (PySpark). | 2 |
|  | Парадигма MapReduce. Сравнение с Hadoop. | 2 |
|  | Параллельная и распределенная обработка больших данных средствами PySpark. | 2 |
|  | Разработка итогового проекта. Постановка задачи организации хранения и обработки данных. Текущий контроль. | 2 |
|  | Проектирование хранилища и процесса обработки данных. | 2 |
|  | Реализация итогового проекта. | 2 |
|  | Презентация результатов итогового проекта. | 2 |
| 2 полугодие 11 класса |
|  | Введение в машинное обучение | Презентация результатов итогового проекта. | 2 |
|  | Типология и метрики качества алгоритмов машинного обучения. | 2 |
|  | Метрические алгоритмы классификации. Текущий контроль. | 2 |
|  | Логические алгоритмы классификации Введение в ансамблевые методы. | 2 |
|  | Модели смесей распределений. Текущий контроль. | 2 |
|  | Методы кластеризации и детектирования аномалий. | 2 |
|  | Методы снижения размерности многомерных данных. | 2 |
|  | Обучение с подкреплением. Текущий контроль. | 2 |
|  | Введение в нейронные сети. | 2 |
|  | Многослойный перцептрон. | 2 |
|  | Свёрточные нейронные сети. Текущий контроль. | 2 |
|  | Рекуррентные нейронные сети. | 2 |
|  | Глубокое обучение без учителя. Текущий контроль. | 2 |
|  | Постановка задачи для итогового проекта. Разработка итогового проекта. | 2 |
|  | Презентация результатов итогового проекта. | 2 |

**ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ИЗУЧЕНИЯ УЧЕБНОГО КУРСА**

Выпускник научится:

* Основами математической статистики и теории вероятностей;
* Находить закономерности в данных, разрабатывать математические модели и модели машинного обучения на эти данных;
* Визуализировать полученные результаты моделирования;
* Основами глубокого обучения.