

УПРАВЛЕНИЕ ОБРАЗОВАНИЯ АДМИНИСТРАЦИИ Г. РЕУТОВ  
Муниципальное бюджетное учреждение дополнительного образования  
«Дом детского творчества»

Московская область, 143966  
г. Реутов, ул. Строителей, д.11

телефон (факс) (495) 528-55-62  
e-mail: info@ddt-reutov.ru

«Согласовано»

Педагогический совет МБУ ДО «ДДТ»

Протокол № 1

от «25» августа 2020 г.



«Утверждено»

Директор МБУ ДО «ДДТ»

Кивва Н.Ю.

Приказ № 31 от 28 августа 2020 г.

**ДОПОЛНИТЕЛЬНАЯ ОБЩЕОБРАЗОВАТЕЛЬНАЯ  
ОБЩЕРАЗВИВАЮЩАЯ ПРОГРАММА  
«DATA SCIENCE. МАШИННОЕ ОБУЧЕНИЕ И НЕЙРОННЫЕ СЕТИ»**

**Направленность:** социально-педагогическая

**Уровень программы:** базовый

**Возраст обучающихся:** 14-17 лет

**Срок реализации:** 1 год

**Автор-составитель:**

Воронич Татьяна Ивановна,  
педагог дополнительного образования

г. Реутов

2020 г.

## ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

На занятиях происходит знакомство обучающихся с одним из самых перспективных направлений нашего времени – машинное обучение. Наука о данных (Data Science) – обширная сфера, которая сочетает в себе несколько смежных дисциплин. Это программирование, математика и статистика, бизнес-аналитика и машинное обучение.

Программа разработана как поэтапный процесс осознания сущности машинного обучения и нейронных сетей в том числе. Акцент в программе сделан не только на совершенствование математического аппарата и навыков программирования (язык Python), но и решения конкретных реальных задач, актуальных в научном мире обработки данных и области технических разработок.

Нормативно-правовые основания:

1. Конвенция о правах ребенка. Принята Генеральной Ассамблеей Организации Объединенных Наций, 20 ноября 1989 г. – ЮНИСЕФ, 1999.
2. Конституция РФ.
3. Федеральный закон от 29 декабря 2012 г. N 273-ФЗ "Об образовании в Российской Федерации".
4. Приказ Министерства просвещения Российской Федерации от 09.11.2018 г. № 196 "Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным общеобразовательным программам".
5. Концепция развития дополнительного образования детей на период до 2020 года включительно (утверждена распоряжением Правительства Российской Федерации от 4 сентября 2014 г. № 1726-р).
6. Методические рекомендации по проектированию дополнительных общеразвивающих программ (включая разноуровневые программы) (Приложение к письму Департамента государственной политики в сфере

воспитания детей и молодежи Министерства образования и науки РФ от 18.11.2015 № 09-3242).

7. Письмо Минобрнауки РФ от 11.12.2006 N 06-1844 "О примерных требованиях к программам дополнительного образования детей".
8. Общие требования к определению нормативных затрат на оказание государственных (муниципальных) услуг в сфере образования, науки и молодежной политики, применяемых при расчете объема субсидии на финансовое обеспечение выполнения государственного (муниципального) задания на оказание государственных (муниципальных) услуг (выполнения работ) государственным (муниципальным) учреждением (утверждены приказом Министерства образования и науки РФ от 22.09.2015 № 1040).
9. Приложение к письму Департамента государственной политики в сфере воспитания детей и молодежи Министерства образования и науки РФ от 14.12.2015 № 09-3564 «О внеурочной деятельности и реализации дополнительных общеобразовательных программ».
10. СанПиН 2.4.4.3172-14 "Санитарно-эпидемиологические требования к устройству, содержанию и организации режима работы образовательных организаций дополнительного образования детей".
11. «Об изучении правил дорожного движения в образовательных учреждениях Московской области».
12. Устав МБУ ДО «Дом детского творчества».

**Направленность** дополнительной образовательной программы «Data Science. Машинное обучение, нейронные сети» - техническая.

**Актуальность программы.** Data Science, или наука о данных, которая включает себя разработки глубоких моделей (нейронных сетей) оказались ключом, подходящим ко всем замкам сразу: новые архитектуры и алгоритмы обучения, а также увеличившиеся вычислительные мощности и появившиеся огромные наборы данных привели к серьезным изменениям в компьютерном зрении, распознавание речи, обработке естественного языка и многое другое.

Спрос на аналитиков данных увеличивается каждый год. А использование в обучающих программах передовые технологии и актуальных концепций современного общества дает возможность ученикам почувствовать себя частью глобального прогресса науки.

***Педагогическая целесообразность.***

***Отличительные особенности данной дополнительной общеобразовательной программы.***

Программы для школьников в данном направлении лишь начинаются появляться в различных кружках и центрах дополнительного образования. Раньше нейронные сети были условной единой областью знаний, сейчас входящие в него подобласти – компьютерное зрение, работа с естественным языком, обучение с подкреплением, генеративно-состязательные сети и другие методы, - выделяются в самостоятельные сферы специализации. В программе «Data Science.Машинное обучение, нейронные сети» подразумевается условное деление на различные подобласти и знакомство с ними. Далее обучающийся выбирает наиболее интересное для себя направление и продолжает заниматься исследованием, проектной деятельностью в данной конкретной области или на стыке областей.

***Цель*** дополнительной образовательной программы «Data Science.Машинное обучение, нейронные сети» - знакомство и создание базы знаний в области анализа данных и возможными инструментами обработки данных, также применение полученных знаний и навыков для последующей проектной деятельности.

***Задачи*** дополнительной образовательной программы «Data Science.Машинное обучение, нейронные сети»:

***Обучающими задачами программы являются:***

- познакомить учеников с тремя основными необходимыми для анализа данных направлениями математики: основы линейной алгебры, математический анализ и теория вероятностей и статистика.

- расширить навыки программирования на языке Python, и познакомить с основными библиотеками для анализа данных;

- обучить ученика собирать, хранить, приводить данные в приемлемый вид – очищать, форматировать и определять, что именно можно узнать из этих данных;

- сформировать способность анализировать данные и обучать алгоритмы, используя математический аппарат, а также подбирать подходящие метрики для оценки качества модели;

***Развивающим задачами программы являются:*** развить у обучающихся логических и аналитических процессов, а также творческой уверенности в создании собственных проектов.

***Воспитательными задачами программы являются:*** воспитать настойчивость и упорство в достижении поставленной цели, создать спокойное отношение к ошибкам и проявление уверенности в собственных силах, воспитать чувства порядочности, аккуратности, честности.

***Адресность*** дополнительной образовательной программы «Data Science. Машинное обучение, нейронные сети» - на программу принимаются все желающие в возрасте 14 до 17 лет.

***Формы и режим занятия*** дополнительной образовательной программы «Data Science. Машинное обучение, нейронные сети». Занятия проводятся 2 раз в неделю по 2 учебных часа. Рекомендуемый состав группы до 12 человек.

***Срок реализации программы*** дополнительной образовательной программы «Data Science. Машинное обучение, нейронные сети» составляет 144 учебных часов.

***Ожидаемые результаты и способы определения результативности программы*** дополнительной образовательной программы «Data Science.Машинное обучение, нейронные сети» являются: получение базовых основ математики (для понимания всего последующего материала) и навыков программирования на языке Python(использование основных библиотек для анализа данных); необходимые знания для построения предсказательных моделей с помощью алгоритмов машинного обучения и оценка их качества; развитие интуитивного понимания работы нейронных сетей, понимание современных архитектур и основных задач, которые можно решать с помощью нейронных сетей. Личностными результатами ожидается приобретение чувства упорства и желания доведения работы до конца, а также систематическая работа над отладкой кода (собственными ошибками). Метапредметными результатами являются проявление творческих способностей обучающихся в создании моделей для решения прикладных задач.

***Формы подведения итогов и реализации программы.*** Промежуточные итоги оцениваются в форме проверочных работ, и проверкой работы кода каждого обучающегося. Итоговой формой результативности программы является создание собственной нейронной сети в одном из направлений - компьютерное зрение или обработка естественного языка на выбор обучающегося.

## СОДЕРЖАНИЕ ПРОГРАММЫ

### УЧЕБНО-ТЕМАТИЧЕСКИЙ ПЛАН ПЕРВЫЙ ГОД ОБУЧЕНИЯ

№	Название раздела, темы	количество часов			Форма контроля
		Всего	Теория	Практика	
1	Вводное занятие.	2	2	0	Диагностика образовательных результатов обучающихся
2	Раздел 1. Знакомство с языком Python.	10	2	8	Демонстрация программы.
3	Раздел 2. Введение в математический анализ.	6	4	2	Письменная проверочная работа.
4	Раздел 3. Основы линейной алгебры.	16	8	8	Демонстрация программы.
5	Раздел 4. Вероятность и случайные величины.	8	4	4	Письменная проверочная работа.
6	Раздел 5. Статистики.	4	2	2	Фронтальный опрос
7	Раздел 6. Машинное обучение и линейные модели.	12	2	10	Демонстрация программы.
8.	Раздел 7. Объектно-ориентированное программирование в Python.	4	2	2	Демонстрация программы.
9	Раздел 8. Введение в нейронные сети.	24	8	16	Демонстрация программы.
10	Раздел 9. Сверточные нейронные сети.	26	6	20	Демонстрация программы.
11	Раздел 10. Рекуррентные нейронные сети.	8	4	4	Демонстрация программы.
12	Раздел 11. Интеллектуальная обработка текстов.	12	4	8	Демонстрация программы.
13	Раздел 12. Современные архитектуры.	12	6	6	Демонстрация программы.
	<b>Всего часов</b>	<b>144</b>	<b>54</b>	<b>90</b>	

# СОДЕРЖАНИЕ

## ПЕРВЫЙ ГОД ОБУЧЕНИЯ

### **Вводное занятие. 2 часа**

Знакомство с содержанием курса и перспективными направлениями в данной области. Инструктаж по технике безопасности.

### **Раздел 1. Знакомство с языком Python.**

#### **Тема 1.1 Основы Python. 2 часа**

*Теория* Ввод данных, основные типы данных, операторы, условный оператор if и тип данных None.

*Практика*

#### **Тема 1.2 Тип переменной «строка». 2 часа**

*Теория* Строки, методы работы со строками.

*Практика*

#### **Тема 1.3 Структуры данных и встроенные функции. – 2 часа.**

*Теория* Знакомство с основными структурами данных. Список. Методы списков. Кортеж.

*Практика*

#### **Тема 1.4 Циклы for и while. – 2 часа.**

*Теория* Итерирование с помощью циклов, особенности построения циклов. Функция range(). Функции enumerate, zip и итерирование по двум аргументам.

*Практика*

#### **Тема 1.5 Структуры данных: множества и словари. – 2 часа.**

*Теория* Множества и методы множеств. Словари, методы и особенности работы со словарем.

*Практика* Решение задач по теме на конкретных данных.

## **Раздел 2. Введение в математический анализ.**

### **Тема 2.1. Функция. – 2 часа.**

*Теория* Функция и ее свойства. Основные функции и графики.

*Практика* Знакомство с библиотекой для визуализации графиков Matplotlib.  
Отрисовка графиков.

### **Тема 2.2. Производная функции. - 2 часа.**

*Теория* Предел и производная. Геометрический и физический смысл производной.

*Практика* Нахождение производных с помощью таблицы производных.  
Знакомство с библиотекой SymPy.

### **Тема 2.3. Производная сложной функции. – 2 часа.**

*Теория* Производная сложной функции. Задача нахождения экстремума.  
Вторая производная и выпуклость.

*Практика* Нахождение производных сложных функций. Решение оптимизационных задач с помощью SciPy.

## **Раздел 3. Основы линейной алгебры.**

### **Тема 3.1. Вектор и операции с ними. - 2 часа.**

*Теория* Векторное пространство. Операции над векторами (сложение, вычитание, умножение). Знакомство с библиотекой NumPy.

*Практика* Задачи, связанные с действиями над векторами. Работа в библиотеке NumPy.

### **Тема 3.2. Матрицы и операции с ними. – 4 часа**

*Теория* Матрица, основные понятия и виды матриц. Операции над матрицами. Матричное произведение. Определитель и ранг матрицы. Транспонирование матриц. Знакомство с библиотекой NumPy для работы с матрицами.

*Практика* Задачи, связанные с действиями над матрицами. Работа в библиотеке NumPy и SciPy.

### **Тема 3.3. Признаковое пространство. – 2 часа**

*Теория* Признаковое описание. Линейная зависимость. Размерность векторного пространства. Нормы: Манхэттенское расстояние и Евклидова норма.

*Практика* Нахождение расстояния (манхэттенское и евклидово) между двумя векторами с помощью библиотеки SciPy. Скалярное произведение и угол между векторами.

### **Тема 3.4. Системы линейных уравнений. – 4 часа**

*Теория* Системы линейных уравнений. Особые виды матриц. Собственные числа и векторы.

*Практика* Основные методы ndarray. Нахождение собственных чисел и векторов матрицы. Пакет numpy.linalg.

### **Тема 3.5. Матричные разложения. – 4 часа**

*Теория* Разложение матриц в произведение. Сингулярное разложение. Приближение матрицей меньшего ранга.

*Практика* Стандартные операции и разложение матриц. Пакет numpy.linalg.

## **Раздел 4. Вероятность и случайные величины.**

### **Тема 4.1. Случайность и вероятность. – 2 часа**

*Теория* Случайность в теории вероятностей и статистики. Вероятность события. Закон больших чисел. Свойства вероятности. Вероятность пары событий. Независимость событий.

*Практика* Задачи по теории вероятностей.

#### **Тема 4.2. Условная вероятность. Теорема Байеса. – 2 часа**

*Теория* Условная и полная вероятность. Формула Байеса. Понятие дискретной случайной величины.

*Практика* Решение задач на определение условной и полной вероятности.

#### **Тема 4.3. Распределения дискретных случайных величин. – 2 часа**

*Теория* Распределение Бернулли. Биномиальное распределение. Элементы комбинаторики: число сочетаний, распределений. Распределение Пуассона.

*Практика* Решение задач с использованием различных законов распределения.

#### **Тема 4.4. Распределения непрерывных случайных величин. – 2 часа**

*Теория* Равномерное распределение. Нормальное распределение.

*Практика* Работа со случайными величинами в ipython notebook. Библиотека random.

### **Раздел 5. Статистики.**

#### **Тема 5.1. Распределения, параметры и оценки. – 2 часа**

*Теория* Оценка распределения по выборке. Важные характеристики распределений. Важные статистики.

*Практика* Статистики с помощью библиотеки statistics.

#### **Тема 5.2. Центральная предельная теорема и доверительные интервалы. – 2 часа**

*Теория* Центральная предельная теорема. Понятие доверительный интервал.

*Практика* Применение ЦПТ для построения доверительного интервала

## **Раздел 6. Машинное обучение и линейные модели.**

### **Тема 6.1. Предобработка данных – 4 часа**

*Теория* Знакомство с машинным обучением. Виды машинного обучения. Обучение с учителем. Признаки в машинном обучении.

*Практика* Знакомство с библиотекой Pandas. Предобработка данных на примере данных «Титаник». Преобразование признаков и статистический анализ данных. Библиотека seaborn.

### **Тема 6.2. Линейные модели в задачах регрессии. – 4 часа**

*Теория* Линейная регрессия. Обучение линейной регрессии. Градиентный спуск для линейной регрессии. Функции потерь. Переобучение. Регуляризация. Метрика качества.

*Практика* Реализация градиентного спуска на примере различных функций. Обучение линейной регрессии. Построение предсказаний. Знакомство с библиотекой Scikit-Learn.

### **Тема 6.3. Линейные модели в задачах классификации. – 4 часа**

*Теория* Логистическая регрессия. Обучение логистической регрессии. Функции потерь в задачах классификации. Метрика качества.

*Практика* Обучение логистической регрессии на данных «Титаник». Использование регуляризации и подбор параметров модели. Знакомство с библиотекой Scikit-Learn.

## **Раздел 7. Объектно-ориентированное программирование в Python.**

## **Тема 7.1. Классы и объекты. – 4 часа**

*Теория* Разделение сущности на классы. Потомки и предки. Определение функций класса. Унаследованные функции. Инициализация объектов.

*Практика* Создание объектов для классов. Использование функции для задания характеристик класса.

## **Раздел 8. Введение в нейронные сети.**

### **Тема 8.1. Граф вычислений и дифференцирование на нем. – 4 часа**

*Теория* Как брать производную на графе вычислений. Алгоритм обратного распространения ошибки.

*Практика* Введение в TensorFlow. Инициализация переменных. Понятие тензор. Обучение линейной регрессии с помощью TensorFlow.

### **Тема 8.2. Перцептрон, нейрон с помощью TensorFlow– 4 часа**

*Теория* Расстояние Кульбака-Лейблера и перекрестная энтропия. Логистический сигмоид и другие функции активации. Обучение глубоких нейронных сетей. Softmax-регрессия.

*Практика* Пример: распознавание рукописных цифр на TensorFlow.

### **Тема 8.3. Перцептрон, нейрон и LogLoss с помощью PyTorch – 4 часа**

*Теория* Введение в PyTorch. Нейрон с разными функциями потерь.

*Практика* Создание и обучение одного нейрона на PyTorch на конкретных датасетах.

### **Тема 8.3. Многослойная сеть на PyTorch– 4 часа**

*Теория* Компоненты нейросети. Forward pass и Backward pass.

*Практика* Создание многослойной нейросети на конкретных датасетах (FashionMnist)

#### **Тема 8.4. Переобучение и регуляризация нейросетей. – 4 часа**

*Теория* Регуляризация в нейронных сетях с помощью дропаут. Инициализация весов. Случайная инициализация и инициализация Ксавье.

*Практика* Задача распознавания рукописных цифр с помощью Keras при разной инициализации весов. Сравнение результатов.

#### **Тема 8.5. Метод Batch Normalization. – 2 часа**

*Теория* Нормализация по мини-батчам. Какие проблемы решает метод и как его использовать.

*Практика* Задача распознавания рукописных цифр с применением нормализации по мини-батчам. Сравнение результатов.

#### **Тема 8.6. Метод моментов и адаптивные варианты градиентного спуска. – 2 часа**

*Теория* Стратегия управления скоростью обучения. Метод моментов. Адаптивные методы оптимизации.

*Практика* Сравнение различных методов усовершенствования обучения нейронных сетей на примере задачи распознавания рукописных цифр. Подведение итогов.

### **Раздел 9. Сверточные нейронные сети.**

#### **Тема 9.1. Введение в сверточные сети и основные понятия. – 4 часа**

*Теория* Карта признаков. Каналы. Ядро свертки. Сверточное преобразование. Слой Pooling. Понятия Padding, stride. Субдискретизация. Receptive field.

*Практика* Операция свертки в TensorFlow. Реализация сверточного слоя и слоя max-pooling с помощью библиотеки NumPy.

### **Тема 9.2. Реализация сверточной нейронной сети– 6 часа**

*Практика* Реализация сверточной нейронной сети на примере задачи классификации картинок на данных CIFAR-10 с помощью TensorFlow и Keras.

Реализация сверточной нейронной сети на примере задачи классификации картинок на данных CIFAR-10 с помощью Pytorch.

### **Тема 9.3. Современные архитектуры нейронных сетей. Transfer learning – 6 часов**

*Теория* VGG, Inception, Resnet. Метод Transfer learning. Обзор моделей VGG, Inception, Resnet в Keras.

*Практика* Реализация сверточной нейронной сети на примере задачи классификации картинок на данных ImageNet с помощью метода Transfer learning на TensorFlow.

### **Тема 9.4. Регуляризация нейронных сетей. Transfer learning – 6 часов**

*Теория* Способы регуляризации нейронных сетей. Augmentation. Transfer learning

*Практика* Реализация сверточной нейронной сети на примере задачи классификации картинок на данных ImageNet с помощью метода Transfer learning на Pytorch. Использование регуляризации.

### **Тема 9.5. Автокодировщики. – 4 часа.**

*Теория* Как работает автокодировщик. Шумоподавляющие и разреженные автокодировщики.

*Практика* Использование автокодировщика на примере датасета рукописных цифр.

## **Раздел 10. Рекуррентные нейронные сети.**

### **Тема 10.1. Обработка последовательностей. – 4 часа.**

*Теория* Типы задач машинного обучения, связанных с последовательностями. Распространение ошибки и архитектуры RNN.

*Практика* Обзор задач с последовательностями.

### **Тема 10.2. Ячейки памяти. – 4 часа.**

*Теория* Двухнаправленные рекуррентные сети. Языковое моделирование. Рекуррентные ячейки памяти: LSTM и GRU.

*Практика* Пример задачи посимвольного порождения текстов с применением ячеек LSTM.

## **Раздел 11. Интеллектуальная обработка текстов.**

### **Тема 11.1. Распределенное представление слов Word2Vec. – 6 часа.**

*Теория* Разновидность модели W2V: CBOW и skip-gram. Понятие embedding. Предобработка данных: токенизация, лемматизация, выделение биграмм/триграмм.

*Практика* Эксперименты с моделями W2V на основании русскоязычной «Википедии». Примеры ближайших соседей и линейных соотношений в модели W2V.

### **Тема 11.2. Задача классификации текстов. – 6 часов.**

*Теория* Постановка задачи. Определение методов решения и способов оценивания.

*Практика* Анализ тональности текстов на данных IMDB (отзывы к фильмам). Предобработка данных и создание собственной нейросети на основе полученных знаний.

## **Раздел 12. Современные архитектуры.**

### **Тема 12.1. Модели с вниманием – 6 часа.**

*Теория* Сети с вниманием. Виды внимания. Машинный перевод. Диалоговая модель.

*Практика* Архитектура encoder-decoder.

### **Тема 12.2. Порождающие модели. – 6 часа.**

*Теория* Таксономия порождающих моделей. Казуальные свертки. Состязательные сети. Дискриминатор и генератор.

*Практика* Трюк с логистическим сигмоидом. Практическое применение порождающей модели.

## **ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ**

Для реализации программы используется помещение, оснащенное ноутбуками для каждого обучающегося, мышки и зарядки. А также используется доска с маркерами и проектор.

Формы организации учебных занятий по программе:

- изложение теоретического материала и решение задач как самостоятельно, так и группой;
- изучение кода программы, проводимыми педагогом в качестве демонстрации;
- самостоятельная реализация кода на своем ноутбуке;

На занятиях предусматриваются следующие формы организации учебной деятельности:

- индивидуальная (работа с раздаточным материалом)
- групповая (обсуждение)

В зависимости от поставленных задач на занятиях используются различные методы и приемы обучения, в том числе:

1) методы стимуляции и мотивации

- вопросы педагога, побуждающие обучающихся к постановке проблемы;
- вопросы, помогающие прояснить ситуацию, выдвинуть гипотезу и понять смысл темы;

2) практические методы:

- групповая реализация кода с обсуждениями каждого шага, а также самостоятельная реализация и работа с ошибками;

### 3)наглядные методы:

- материал представляется объяснением педагога у доски, а также в виде презентации;
- используются различные наглядные материалы в виде таблиц, графиков и схем.

#### Алгоритм проведения занятия:

- вопросы по предыдущему материалу, которые могут быть связаны с новой темой и обсуждение их.
- знакомство с новой темой в теоретическом виде;
- реализация заданий новой темы на ноутбуках;
- подведение итогов работы на занятии.

#### Контроль и оценка знаний обучающихся:

Контроль обучающихся проводится в результате проверки (промежуточной, итоговой) в форме фронтальных и индивидуальных устных опросов, практических заданий. Фронтальный опрос проводится как беседа, в котором участвуют все обучающиеся. Педагог подготавливает серию вопросов по конкретной теме курса, на которые обучающиеся дают короткие обоснованные ответы.

#### Список используемой литературы

1. Николенко С., Кадурич А., Архангельская Е. Глубокое обучение – СПб.: Питер, 2018 – 480 с.
2. Курс «Прикладные задачи анализа данных». Автор: Moscow Institute of Physics and Technology, Yandex <https://ru.coursera.org/lecture/data-analysis-applications/analiz-tonal-nosti-tieksta-r79kd>
3. Открытый курс машинного обучения/ Open Machine Learning Course mlcourse.ai. <https://habr.com/company/ods/blog/323890/>

### Список литературы для обучающихся и родителей

1. Николенко С., Кадури́н А., Архангельская Е. Глубокое обучение – СПб.: Питер, 2018 – 480 с.