

*Ассоциация участников технологических кружков*

УТВЕРЖДАЮ

Президент Ассоциации



А.И. Федосеев

Дополнительная общеразвивающая образовательная программа

**«Машинное обучение»**

Профиль: «Искусственный интеллект»

общая трудоемкость – 86 часов

образовательные технологии – предметно-ориентированные

направленность: техническая

**Разработчики:**

команда разработчиков профиля  
«Искусственный интеллект» НТО

Ассоциация участников  
технологических кружков

Москва, 2021 г.

## Пояснительная записка

### Анонс программы

Данная программа объединила в себе лидеров рынка по направлению Data Science и исследователей данных Сбербанка, чтобы каждый ученик мог прокачать свои возможности от 0 до уровня, достаточного для участия в олимпиадах и соревнованиях.

Мир в котором мы живем невозможно представить без информационных технологий. С каждым годом они стремительно развиваются и делают жизнь человека более комфортной. Особое место среди IT разработок занимает искусственный интеллект. Он постепенно проникает во многие области жизни человека. То что можно создать в сфере искусственного интеллекта интересно всем. Искусственный интеллект обогатит в любом случае любую специальность. Какой бы опыт ученики не имели перед началом изучения данного курса их ждет насыщенная программа обучения, возможно местами сложная, но в целом очень интересная. Все знания полученные в рамках данного курса позволят сформировать у учащихся навыки XXI века.

В основе курса лежит практико-ориентированный подход, который позволяет разрабатывать проекты, которые могут что-нибудь изменить, которые могут быть полезны на практике.

Цель курса: предоставить учащимся возможность осознанно подойти к выбору карьеры в IT-индустрии и поступлению в вуз на специальность, связанную с программированием или машинным обучением.

### Задачи курса:

- познакомить учащихся с понятием искусственного интеллекта и его возможностями;
- сформировать у учащихся базовые знания о языке Python;
- познакомить учащихся с объектно-ориентированным программированием;
- научить учащихся работать с моделями машинного обучения.

Категория учащихся: курс рассчитан на учащихся 8-11 классов

Сроки реализации программы: курс рассчитан на 86 часов (5 месяцев занятий по 4 часа в неделю)

По итогам обучения на курсе учащиеся смогут:

- принимать участие в олимпиадах по искусственному интеллекту
- принимать участие в международных соревнованиях по искусственному интеллекту
- принять участие в стажировках для школьников

Курс состоит из тематических уроков, а также включает в себя несколько проектных заданий.

Все тематические уроки разбиты на семь модулей:

- Основы Python.
- Анализ данных.
- Машинное обучение.

- Решение задачи классификации
- Рекомендательные системы
- Нейросети
- Обработка естественного языка

Рекомендуется следующая структура любого урока:

- вводный блок;
- постановка задачи;
- теория;
- разбор примера вместе с учащимися;
- отработка практики самостоятельно;
- подведение итогов урока.

По итогам курса ученики:

- сразу начнут писать программы на языке Python и осознанно использовать различные методы изученных библиотек;
- узнают различные алгоритмы и методы машинного обучения и смогут применять их на практике;
- разработают предсказательную модель ML.

### Планируемые результаты программы

Предметные результаты обучения:

- иметь представление об искусственном интеллекте и его возможностях в современном обществе;
- иметь представление о перспективах развития технологий искусственного интеллекта;
- уметь приводить примеры решения различных задач (разными методами) с использованием систем искусственного интеллекта;
- использовать различные средства для поиска информации о возможностях, направлениях развития искусственного интеллекта и его средств;
- уметь взаимодействовать со средствами искусственного интеллекта;
- иметь представление о возможностях и перспективах машинного обучения систем искусственного интеллекта, об основных подходах;
- приводить примеры типовых задач;
- иметь представление о проблемах машинного обучения;
- знать основы машинного обучения, в том числе математические;
- иметь представление о различных подходах к моделированию интеллектуальной деятельности;
- приводить примеры решения различных задач с использованием нейронных сетей;
- иметь представление о перспективных направлениях моделирования интеллектуальной деятельности;
- уметь обучать готовые нейронные сети при проведении компьютерных экспериментов;
- иметь представление о возможностях и перспективах развития искусственного интеллекта в области обработки естественного языка интеллектуальными системами;

- приводить примеры решения различных задач с использованием систем искусственного интеллекта при распознавании и синтезе речи;
- иметь представление о проблемах обучения интеллектуальных систем распознаванию и синтезу речи;
- уметь разрабатывать некоторые программные средства по обработке естественного языка.

Метапредметные результаты обучения:

- формировать универсальные учебные действия (познавательные, регулятивные, коммуникативные), обобщенные способы информационной деятельности при использовании информационных технологий, в том числе при организации машинного обучения интеллектуальных систем, при моделировании интеллектуальной деятельности и при использовании технологий искусственного интеллекта по обработке естественного языка;
- развить познавательные интересы, интеллектуальные и творческие способности путем освоения и использования методов и средств технологий искусственного интеллекта, при проведении компьютерных экспериментов по машинному обучению, а также моделировании интеллектуальной деятельности;
- приобрести опыт машинного обучения интеллектуальных систем и обучения нейронных сетей в индивидуальной, групповой и коллективной учебно-познавательной деятельности.

Личностные результаты обучения:

- личностное и предпрофессиональное самоопределение через познавательную мотивацию к получению профессий, связанных с искусственным интеллектом и через познавательный интерес к достижениям в области машинного обучения и распознавания и синтеза речи;
- построение дальнейшей индивидуальной образовательной траектории через получение представления о перспективных направлениях машинного обучения интеллектуальных систем;
- осознание стратегической важности для государства, общества и для своего будущего развития машинного обучения интеллектуальных систем.

## Содержание программы

### Учебно-тематический план

№	Тема	Количество часов	Лекции	Практика	Контроль
Вводное занятие					
1	Введение в искусственный интеллект	0, 3	0,2	0,1	
2	Введение в Python	0, 7	0,2	0,5	
	Всего по разделу	1	0,4	0,6	
Основы Python					

3	Переменные и условный оператор	3	1	2	Тест
4	Циклы и работа со строками	3	1	2	Тест
5	Списки и словари	4	1	3	Тест
6	Работа с файлами	2	0,6	1,4	Тест
7	Функции	3	0,5	2,5	Тест
8	Основы объектно-ориентированного программирования	3	0,5	2,5	Тест
	Всего по разделу	18	4,6	13,4	
<b>Анализ данных</b>					
9	Введение в анализ данных	5	1,5	3,5	Тест
10	Работа с таблицами и подготовка данных	4	1	3	Тест
11	Визуализация данных	5	1,5	3,5	Тест
12	Линейная алгебра с NumPy	4	1	3	Тест
13	Статистика и теория вероятности	5	1,5	3,5	Тест
	Всего по разделу	23	6,5	16,5	
<b>Машинное обучение</b>					
14	Введение в машинное обучение	5	2	3	Тест
15	Линейная регрессия	4	1,5	2,5	Тест
16	Логистическая регрессия	4	1,5	2,5	Тест
17	Дерево решений и случайный лес	4	1,5	2,5	Тест
18	Ансамбли	4	1,5	2,5	Тест
19	Определение класса игрока Dota 2. Проект	2		2	Проект
	Всего по разделу	23	8	15	
<b>Решение задачи классификации</b>					
20	Предсказание транзакций. Проект	2		2	Проект
	Всего по разделу	2		2	
<b>Рекомендательные системы</b>					
21	Введение в рекомендательные системы	2	1,5	0,5	Тест
22	Рейтинговая рекомендательная система. Проект	2	0,5	1,5	Проект

23	Контентная рекомендательная система. Проект	2	0,5	1,5	Проект
24	Коллаборативная рекомендательная система. Проект	3	0,5	2,5	Проект
	Всего по разделу	9	3	6	
Нейросети					
25	Введение в нейросети	2	1,5	0,5	Тест
26	Нейросети. Проект	2		2	Проект
	Всего по разделу	4	1,5	2,5	
Обработка естественного языка					
27	Введение в NLP	2	1,5	0,5	Тест
28	Вопросно-ответные системы. Проект	3	0,5	2,5	Проект
	Всего по разделу	5	2	3	
Итоговое занятие					
29	Подведение итогов курса	1		1	
	Всего по разделу	1		1	
	Итого на курс	86	26	60	

## Содержание программы

Наименование, разделов, тем	Содержание	Виды учебных занятий, учебных работ
Раздел 1 Вводное занятие (1 ч)		
Введение в искусственный интеллект	Понятие искусственного интеллекта (ИИ). Области применения. Обсуждение, где встречаемся с ИИ, перспективы развития	Лекция обсуждение
Введение в Python	Технические моменты: установка Anaconda.	Лекция практика

Раздел 2 Основы Python (18 ч)		
Переменные и условный оператор	Описание переменных. Операции над переменными. Условный оператор. Полный условный оператор.	Лекция Тест №№ Задание №№
Циклы и работа со строками	Цикл While. Цикл For. Операции над строками.	Лекция Тест №№ Задание №№
Списки и словари	Списки. Индексы и срезы. Методы работы со списками. Генератор списков. Словари. Перебор элементов словаря.	Лекция Тест №№ Задание №№
Работа с файлами	Функция open(). Считывание данных в программу из внешнего файла. Базовые операции для анализа полученных данных.	Лекция Тест №№ Задание №№
Функции	Определение функции. Лямбда-функция и функция map. Локальные и глобальные переменные.	Лекция Тест №№ Задание №№
Основы объектно-ориентированного программирования	Основные понятия ООП. Создание класса. Создание экземпляра класса. Инкапсуляция. Наследование и полиморфизм. Модули.	Лекция Тест №№ Задание №№
Раздел 3 Анализ данных (23 ч)		
Введение в анализ данных	О Pandas. Основные объекты Pandas. Считывание и вывод данных. Информация о данных. Извлечение данных по условиям. Группировка данных. Агрегирующие функции.	Лекция Тест №№ Задание №№
Работа с таблицами и подготовка данных	Сводные таблицы. Вводные данные. Объединение таблиц. Предварительная обработка данных.	Лекция Тест №№ Задание №№
Визуализация данных	Графические возможности Pandas. Объектно-ориентированный подход. Гистограммы распределения признаков	Лекция Тест №№ Задание №№

	в Seaborn. Работа с категориальными признаками. Библиотека Plotly.	
Линейная алгебра с NumPy	Массивы NumPy. Знакомство с линейной алгеброй. Операции с массивами.	Лекция Тест №№ Задание №№
Статистика и теория вероятности	Введение в статистику и теорию вероятности. Статистические операции в NumPy. Графики табличного типа. Распределение данных.	Лекция Тест №№ Задание №№
Раздел 4 Машинное обучение (23 ч)		
Введение в машинное обучение	Задачи машинного обучения. Этапы разработки модели. Проблема переобучения. Библиотека Scikit-learn.	Лекция Тест №№ Задание №№
Линейная регрессия	Описание линейной регрессии. Метрики качества линейной регрессии. Разбор задачи линейной регрессии.	Лекция Тест №№ Задание №№
Логистическая регрессия	Логистическая регрессия. Разбор задачи логистической регрессии. Метрики качества классификации. Пример использования метрик качества в решении задачи. Решение задачи логистической регрессии на kaggle.	Лекция Тест №№ Задание №№
Дерево решений и случайный лес	Дерево решений. Реализация модели дерева решений для задач бинарной классификации. Решение задачи деревом решений на kaggle. Случайный лес реализация модели случайного леса для решения задачи регрессии. Решение задачи случайным лесом на kaggle.	Лекция Тест №№ Задание №№
Ансамбли	Бустинг. Реализация градиентного бустинга в задаче классификации. Стекинг. Реализация блендинга в задаче классификации. Решение задачи моделями бустинга и стекинга на kaggle.	Лекция Тест №№ Задание №№
Определение класса игрока Dota 2. Проект	Решение задачи классификации на данных игры Dota 2.	Проект

Раздел 5 Решение задачи классификации (2 ч)		
Предсказание транзакций. Проект	Решение задачи по предсказанию вероятности совершения покупки клиентом банка.	Проект
Раздел 6 Рекомендательные системы. Проект (9 ч)		
Введение в рекомендательные системы	Введение в рекомендательные системы. Что такое рекомендательная система? Неперсонализированные рекомендации. Контентные рекомендации. Коллаборативная фильтрация.	Лекция Тест
Рейтинговая рекомендательная система. Проект	Рейтинговая рекомендательная система.	Лекция Проект
Контентная рекомендательная система. Проект	Контентная рекомендательная система.	Лекция Проект
Коллаборативная рекомендательная система. Проект	Коллаборативная рекомендательная система.	Лекция Проект
Раздел 7 Нейросети (4 ч)		
Введение в нейросети	Знакомство с нейросетями. Задачи обучения нейросетей.	Лекция Тест
Нейросети. Проект	Разработка нейросети: решение проектной задачи.	Лекция Проект
Раздел 8 Обработка естественного языка (5 ч)		
Введение в NLP	Анализ элементов естественного языка. Методы обработки естественного языка. Применение методов обработки естественного языка.	Лекция Тест

Вопросно-ответные системы. Проект	Разработка вопросно-ответной системы: решение проектной задачи.	Лекция Проект
Раздел 9 Итоговое занятие (1 ч)		
Подведение итогов курса	Теоретические и прикладные результаты изучения разделов курса.	Обсуждение

## Результаты обучения

### Раздел 1 Вводное занятие (1 ч)

Введение в искусственный интеллект. Введение в Python

Учащиеся должны знать:

- понятие искусственного интеллекта

Учащиеся должны понимать:

- возможности применения искусственного интеллекта в различных областях жизни

Учащиеся должны уметь:

- устанавливать программного обеспечения, необходимого для программирования на Python

### Раздел 2 Основы Python (18 ч)

#### Переменные и условный оператор

Учащиеся должны знать:

- что такое переменные;
- основные типы переменных - int, float, str и bool;
- две новые функции input() и print();
- что такое условный оператор.

Учащиеся должны понимать:

- как писать свои программы в Jupiter;
- как запускать программы;
- для чего нужна функция print();
- для чего нужна функция input();
- как использовать условный оператор для составления ветвления в программе.

Учащиеся должны уметь:

- читать простые программы, включающие в себя переменные, операции над ними, условные операторы, ввод данных с клавиатуры и вывод информации на экран;



- работать с переменными при написании своих программ;
- составлять логические условия, которые требуются для решения поставленной задачи;
- использовать неполный условный оператор для проверки одного условия;
- использовать полный условный оператор в своих программах для решения задач с несколькими условиями.

### **Циклы и работа со строками**

Учащиеся должны знать:

- что такое цикл;
- структуры циклов while и for;
- операторы break и continue;
- функции работы со строками: len(), find(), count(), lower(), upper(), title(), replace().

Учащиеся должны понимать:

- как использовать циклы для использования повторений в программе;
- для каких задач удобно использовать цикл while, а для каких for;
- для чего нужна каждая из рассмотренных функций по работе со строками;
- как использовать циклы для организаций повторений в программе.

Учащиеся должны уметь:

- читать программы, включающие в себя циклы, а также работу со строками;
- работать с циклами и строками при написании своих программ;
- составлять условия в заголовке цикла while, которые требуются для решения поставленной задачи.

### **Списки и словари**

Учащиеся должны знать:

- что такое список;
- что такое генератор списка;
- что такое словарь.

Учащиеся должны понимать:

- как создать список;
- как создать список при помощи генератора списка;
- как добавить или изменить элементы списка;
- как создать словарь;
- как добавить или изменить элементы словаря.

Учащиеся должны уметь:

- создавать списки и словари;
- добавлять элементы в список или словарь;
- изменять элементы списка или словаря;
- обращаться к конкретному элементу списка или словаря;
- выводить списки или их отдельные элементы на экран;
- выводить словари целиком, ключи или значения конкретных элементов словаря на экран.

### **Работа с файлами**

Учащиеся должны знать:

- команду, позволяющую совершать манипуляции с файлами csv в программе.

Учащиеся должны понимать:

- как открыть файл для работы с ним в программе;
- почему нужно закрывать файл после работы с ним в программе.

Учащиеся должны уметь:

- считывать данные из файла в различные структуры в программе;
- анализировать данные, полученные из внешнего файла csv.

### **Функции**

Учащиеся должны знать:

- что такое функция;
- что такое анонимная функция;
- что такое локальные и глобальные переменные.

Учащиеся должны понимать:

- для чего нужны функции;
- как вызвать функции;
- как применить функцию сразу ко всему списку при помощи функции map;
- где у нас в программе находятся глобальные переменные, а где локальные.

Учащиеся должны уметь:

- создавать (определять) свои функции;
- вызывать функции для их использования в программе.

### **Основы объектно-ориентированного программирования**

Учащиеся должны знать:

- что такое класс;
- что такое экземпляр класса;
- что такое полиморфизм;
- что такое инкапсуляция;
- что такое наследование.

Учащиеся должны понимать:

- как создавать свои классы;
- чем методы отличаются от обычных функций;
- чем свойства отличаются от полей;
- для чего нужны модули.

Учащиеся должны уметь:

- создавать свои классы и их экземпляры;
- подключать необходимые для работы модули или функции, хранящиеся в них.

## Раздел 3. Анализ данных (23 ч)

### **Введение в анализ данных**

Учащиеся должны знать:

- что такое группировка данных;
- основные объекты в Pandas — Series и DataFrame;
- что такое группировка данных;

- что такое сводная таблица.

Учащиеся должны понимать:

- как подключить библиотеку для анализа данных Pandas;
- как вывести на экран часть загруженных данных;
- как вывести на экран информацию о загруженных данных;
- зачем нужна группировка данных;
- для чего нужны сводные таблицы.

Учащиеся должны уметь:

- выводить на экран загруженные данные;
- извлекать данные из датасета по определенным условиям;
- группировать данные по различным признакам;
- создавать сводные таблицы на основе датасетов.

### **Работа с таблицами и подготовка данных**

Учащиеся должны знать:

- метод, позволяющий объединять таблицы;
- какие трудности возникают при объединении данных;
- что такое дубликаты.

Учащиеся должны понимать:

- зачем хранить информацию в разных таблицах;
- какие есть варианты объединения таблиц и как они отличаются;
- как удалять повторы строк;
- как удалять пустые значения;
- зачем стоит заменять пустые значения на константы;
- как отфильтровать значения различными функциями.

Учащиеся должны уметь:

- объединять таблицы;
- удалять дубликаты;
- фильтровать данные;
- удалять пропуски;
- заменять пропуски.

### **Визуализация данных**

Учащиеся должны знать:

- о встроенных возможностях визуализации библиотеки Pandas;
- диаграммы и графики, которые можно построить в библиотеке Matplotlib;
- диаграммы и графики, которые можно построить в библиотеке Seaborn;
- диаграммы и графики, которые можно построить в библиотеке Plotly.

Учащиеся должны понимать:

- как подключать библиотеки для создания графиков;
- чем отличаются виды графиков и диаграмм;
- для каких целей подходят разные библиотеки для визуализации данных.

Учащиеся должны уметь:

- строить графики, используя библиотеки Pandas, Matplotlib, Seaborn и Plotly;

- настраивать параметры построенных графиков.

### **Линейная алгебра с NumPy**

Учащиеся должны знать:

- что такое массив и его особенности;
- что такое вектор и матрица.

Учащиеся должны понимать:

- зачем нужны массивы;
- чем отличается массив от списка list;
- как обратиться к отдельным элементам массива.

Учащиеся должны уметь:

- создавать массивы;
- работать с индексами в массивах;
- совершать различные операции с массивами и их элементами;
- совершать математические операции с векторами и матрицами.

### **Статистика и теория вероятности**

Учащиеся должны знать:

- что такое нормальное распределение;
- как выглядит «ящик с усами»;
- что такое квантили и выбросы.

Учащиеся должны понимать:

- как выполняются статистические операции;
- что отображает нормальное распределение.

Учащиеся должны уметь:

- строить диаграмму «ящик с усами» и использовать ее в анализе;
- считать статистические характеристики в Python и визуализировать их.

## Раздел 4 Машинное обучение (23 ч)

### **Введение в машинное обучение**

Учащиеся должны знать:

- что такое обучение с учителем;
- что такое обучение без учителя;
- что такое кросс-валидация;
- как проходит процесс разработки модели машинного обучения.

Учащиеся должны понимать:

- проблемы, которые могут возникнуть при обучении модели;
- способы борьбы с проблемой переобучения.

Учащиеся должны уметь:

- выбирать подходящий алгоритм для разных задач машинного обучения;

- подключать библиотеку Scikit-learn.

### **Линейная регрессия**

Учащиеся должны знать:

- что такое линейная регрессия;
- как решается задача линейной регрессии;
- некоторые метрики качества модели.

Учащиеся должны понимать:

- требования, которым должны соответствовать данные при работе с линейной регрессией;
- процесс создания модели линейной регрессии;
- порядок решения задач линейной регрессии при помощи библиотеки Scikit-learn.

Учащиеся должны уметь:

- находить ошибку модели линейной регрессии;
- решать задачу линейной регрессии с библиотекой Scikit-learn.

### **Логистическая регрессия**

Учащиеся должны знать:

- что такое логистическая регрессия;
- что является решением задачи логистической регрессии;
- некоторые метрики качества классификации;
- что такое матрица ошибок.

учащиеся должны понимать:

- задачу линейной классификации;
- как строится модель логистической регрессии;
- как оценить качество полученного результата;
- как решить задачу линейной классификации.

Учащиеся должны уметь:

- использовать метрики качества классификации для оценки модели;
- решать задачу логистической регрессии методами библиотеки Scikit-learn.

### **Дерево решений и случайный лес**

Учащиеся должны знать:

- что такое дерево решений;
- что такое случайный лес;
- как решать задачи регрессии и классификации с помощью случайного леса;
- метрики качества дерева решений и случайного леса.

Учащиеся должны понимать:

- почему дерево решений является популярным алгоритмом;
- как случайный лес улучшает работу метода дерева решений;
- как строится модель дерева решений;

- как строится модель случайного леса;
- как оценить качество полученного результата.

Учащиеся должны уметь:

- строить дерево решений и случайный лес, используя библиотеку Scikit-learn;
- решать задачи классификации и регрессии при помощи дерева решений или случайного леса методами библиотеки Scikit-learn.

### **Ансамбли**

Учащиеся должны знать:

- что такое бустинг и бэггинг;
- что такое стекинг и блендинг.

Учащиеся должны понимать:

- зачем нужны ансамбли;
- как строятся ансамбли;
- как решить задачу с помощью ансамблей.

Учащиеся должны уметь:

- использовать новые знания на практике;
- решать задачи методами библиотеки sklearn.

## Раздел 6 Рекомендательные системы. Проект (9 ч)

### **Введение в рекомендательные системы**

Ученик должен знать:

- множество сфер применения рекомендательных систем с примерами;

Ученик должен понимать

- принципы рекомендательных систем и их виды;

Ученик может уметь:

- объяснить матрицу предпочтений;
- описать три подхода к созданию систем рекомендаций;
- создавать рекомендательные системы на основе трех базовых подходов.