

МИНИСТЕРСТВО ОБРАЗОВАНИЯ МОСКОВСКОЙ ОБЛАСТИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»
АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБЩЕОБРАЗОВАТЕЛЬНАЯ ОРГАНИЗАЦИЯ
«ФИЗТЕХ-ЛИЦЕЙ» ИМЕНИ П. Л. КАПИЦЫ

УТВЕРЖДАЮ
Директор АНОО "Физтех-лицей" им. П.Л. Капицы



М.Г. Машкова
Приказ №
от "4" * сент. 2025 г.

ПРИНЯТО

на заседании
методического совета

Протокол № 3
от 19.06.2025

СОГЛАСОВАНО

Руководитель кафедры
словесности

Никитенко Л. В..

СОГЛАСОВАНО

заместитель директора по
дополнительному
образованию

Меньшакова С. А.

ДОПОЛНИТЕЛЬНАЯ ОБЩЕОБРАЗОВАТЕЛЬНАЯ ОБЩЕРАЗВИВАЮЩАЯ
ПРОГРАММА

«Машинное обучение и нейросети в задачах обработки естественного языка»
(гуманитарной направленности)

Уровень программы: базовый
Срок реализации программы: 7 месяцев
Возрастная категория 16-17 лет
Состав группы: 10 -15 обучающихся
Форма обучения: очная

Составитель: Дедов Т. Г.
педагог дополнительного образования

г. Долгопрудный, 2025 г.

СОДЕРЖАНИЕ

Введение

Нормативно-правовая база

Раздел 1. «Комплекс основных характеристик образования»

Пояснительная записка

Цели и задачи

Содержание программы

Планируемые результаты

Раздел 2. «Комплекс организационно-педагогических условий»

Календарный учебный график

Условия реализации программы

Формы аттестации

Оценочные материалы

Методические материалы

Раздел о воспитании

Список литературы

Приложения

1. РАЗДЕЛ 1. «КОМПЛЕКС ОСНОВНЫХ ХАРАКТЕРИСТИК ОБРАЗОВАНИЯ»

1.1. Пояснительная записка

- Направленность программы: естественнонаучная.
- Актуальность программы вызвана стремительным развитием технологий искусственного интеллекта и повсеместным внедрением систем машинного обучения, особенно в области обработки естественного языка (NLP). Навыки работы с данными, программирования и понимания основ нейросетей являются критически важными для будущих специалистов в IT, лингвистике, социальных и гуманитарных науках.
- Педагогическая целесообразность заключается в предоставлении обучающимся возможности углубленно изучить перспективную и востребованную область на стыке компьютерных наук и лингвистики, развить алгоритмическое и аналитическое мышление, а также получить практические навыки для реализации собственных проектов.
- Адресат программы: обучающиеся 10-11 классов (16-17 лет), проявляющие интерес к программированию, математике, компьютерным технологиям и лингвистике. Программа может быть адаптирована для мотивированных учащихся.
- Уровень программы: базовый.
- Объем и срок реализации: общий объем программы — 56 академических часов. Срок реализации — 7 месяцев.
- Формы обучения: очная.
- Режим занятий: 2 академических часа 1 раз в неделю.
- Особенности организации образовательного процесса: программа сочетает лекционные занятия, практикумы, проектную деятельность и самостоятельную работу. Основной упор делается на решение прикладных задач и выполнение проектов.

1.2. Цель и задачи

- Цель: формирование у обучающихся систематических знаний и практических умений в области машинного обучения и обработки естественного языка, развитие способности к самостоятельному решению прикладных задач с использованием современных IT-технологий.
- Задачи:
 - Предметные:
 - Сформировать навыки программирования на Python для анализа данных.
 - Обучить основам работы с ключевыми библиотеками для Data Science (NumPy, Pandas, Matplotlib/Seaborn, Scikit-learn).
 - Раскрыть основные концепции и методы машинного обучения (классификация, регрессия, кластеризация).
 - Дать представление об архитектуре и принципах работы нейросетей.

- Сформировать умения применять методы машинного обучения для обработки текстовой информации (парсинг, векторизация, создание чат-ботов).
- Метапредметные:
 - Развивать навыки самостоятельного планирования и реализации учебных и проектных задач.
 - Формировать умения критически оценивать информацию, выбирать оптимальные методы решения задач.
 - Развивать коммуникативные навыки через презентацию и защиту собственных проектов.
 - Личностные:
 - Воспитывать целеустремленность, ответственность и трудолюбие.
 - Формировать готовность к самообразованию и непрерывному обучению.
 - Развивать креативность и инициативность в проектной деятельности.

1.3. Содержание программы

Содержание программы структурировано по темам в соответствии с учебно-тематическим планом и включает теоретическую и практическую составляющие.

№ п/п	Наименование раздела, темы	Количество часов			Формы аттестации/ контроля
		Всего	Теория	Практика	
1	Раздел 1. Введение в инструментарий	8	2	6	
1.1.	Введение. Цели и задачи курса. Повторение Python.	2	1	1	Устный опрос

1.2.	Библиотека NumPy для научных вычислений.	2	1	1	Решение задач
1.3.	Библиотека Pandas для анализа данных.	2	1	1	Решение задач
1.4.	Визуализация данных с Matplotlib и Seaborn.	2	1	1	Решение задач
2	Раздел 2. Основы машинного обучения	22	6	16	
2.1.	Машинное обучение: основы и задачи.	2	1	1	Устный опрос
2.2.	Машинное обучение: задачи классификации.	4	1	3	Решение задач

2.3.	Машинное обучение: задачи кластеризации.	4	1	3	Решение задач
2.4.	Машинное обучение: линейная регрессия.	4	1	3	Решение задач
2.5.	Другие методы: деревья решений.	4	1	3	Решение задач
2.6.	Отбор и ранжирование признаков.	2	1	1	Устный опрос
2.7.	Функции потерь и метрики качества.	2	1	1	Устный опрос
3	Раздел 3. Продвинутые методы и работа с текстом	18	4	14	
3.1.	Композиция моделей.	2	0.5	1.5	Решение задач

3.2.	Градиентный бустинг.	4	1	3	Тестовая работа
3.3.	Работа с текстами: парсинг сайтов.	6	1	5	Решение задач
3.4.	Работа с текстами: создание Telegram-ботов.	6	1	5	Решение задач
4	Раздел 4. Введение в глубинное обучение	6	2	4	
4.1.	Введение в глубинное обучение.	2	1	1	Устный опрос
4.2.	Полносвязные и сверточные слои.	4	1	3	Устный опрос, Тест
5	Раздел 5. Итоговый проект	2	0	2	

5.1.	Презентация собственных исследований.	2	0	2	Защита проекта
	ИТОГО:	56	14	42	

Содержание разделов и тем:

Раздел 1. Введение в инструментарий

- Тема 1.1. Знакомство с целями и структурой курса. Актуализация знаний Python: базовые структуры данных, функции, основные конструкции. Знакомство со средой разработки (Jupyter Notebook).
- Тема 1.2. Библиотека NumPy: создание массивов, основные операции, индексация, линейная алгебра. Практическое применение для математических вычислений.
- Тема 1.3. Библиотека Pandas: структуры Series и DataFrame, загрузка данных из CSV-файлов, базовые операции (фильтрация, агрегация, группировка).
- Тема 1.4. Построение графиков с помощью Matplotlib и Seaborn: гистограммы, scatter plots, box plots. Инструменты для визуального анализа данных.

Раздел 2. Основы машинного обучения

- Тема 2.1. Основные понятия: обучение с учителем и без учителя, переобучение/недообучение, разбиение данных на выборки (train/test). Базовые алгоритмы библиотеки Scikit-learn.
- Тема 2.2. Задачи классификации: логистическая регрессия, метод k-ближайших соседей (KNN), метод опорных векторов (SVM). Оценка качества моделей.
- Тема 2.3. Задачи кластеризации: алгоритм K-средних (K-Means). Поиск паттернов в данных, интерпретация результатов.
- Тема 2.4. Линейная регрессия: математическое обоснование, градиентный спуск, оценка точности предсказания.
- Тема 2.5. Деревья решений: принцип работы, построение, преимущества и недостатки.
- Тема 2.6. Методы отбора признаков (Feature Selection), анализ важности признаков для улучшения качества модели.
- Тема 2.7. Функции потерь (MSE, Log Loss), метрики качества (Accuracy, Precision, Recall, F1-score).

Раздел 3. Продвинутое методы и работа с текстом

- Тема 3.1. Ансамбли моделей: бэггинг и случайный лес.
- Тема 3.2. Градиентный бустинг: принцип работы, обзор библиотек (XGBoost, CatBoost).
- Тема 3.3. Техники сбора текстовых данных: веб-скрейпинг с помощью библиотек BeautifulSoup/Scrapy. Очистка и предобработка текста (токенизация, лемматизация).
- Тема 3.4. Векторизация текста (TF-IDF, Word2Vec). Создание простого чат-бота для мессенджера Telegram с использованием полученных навыков.

Раздел 4. Введение в глубинное обучение

- Тема 4.1. От машинного обучения к глубинному. Искусственные нейроны, архитектура полносвязной нейронной сети, прямой проход.
- Тема 4.2. Принцип работы сверточных слоев. Обзор фреймворка Keras/TensorFlow/PyTorch для построения простых нейросетей.

Раздел 5. Итоговый проект

- Тема 5.1. Самостоятельная разработка и реализация небольшого исследовательского проекта по анализу текстовых данных или созданию ML-модели. Подготовка презентации и защита результатов перед группой.

1.4. Планируемые результаты

Личностные:

- Сформированность мотивации к саморазвитию и непрерывному обучению в области ИТ.
- Проявление инициативности и креативности в проектной деятельности.
- Готовность к сотрудничеству при работе над групповыми проектами.

Метапредметные:

- Умение самостоятельно ставить цели, планировать и корректировать свою деятельность для достижения результата.
- Владение навыками поиска, анализа и критической оценки информации из различных источников.
- Способность ясно и логично излагать свои мысли, аргументировать свою точку зрения при защите проекта.

Предметные:

- Знание и умение применять на практике основные библиотеки Python для анализа данных (NumPy, Pandas, Matplotlib/Seaborn, Scikit-learn).
- Понимание и способность реализовывать базовые алгоритмы машинного обучения для задач классификации, регрессии и кластеризации.
- Умение собирать, обрабатывать и анализировать текстовые данные, включая создание простых чат-ботов.

- Понимание базовых принципов работы нейронных сетей.
- Способность провести небольшое самостоятельное исследование или реализовать проект с использованием изученных технологий.

2. РАЗДЕЛ 2. «КОМПЛЕКС ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИХ УСЛОВИЙ»

2.1. Календарный учебный график

№ п/п	Дата пров еден ия	Тема занятия	Кол -во час ов	Форма занятия	Место проведени я	Форм а контр оля
1	04.1 0.20 25	Введение. Цели и задачи изучаемого курса. Повторение некоторых инструментов языка Python	2	Лекция с элементами практикума	Кабинет информати ки №301	Устный опрос
2	11.1 0.20 25	Библиотека NumPy и ее важность для задач машинного обучения	2	Практикум	Кабинет информати ки №301	Решение задач

3	18.1 0.20 25	Библиотека Pandas и ее использование для анализа данных	2	Практикум	Кабинет информати ки №301	Решение задач
4	25.1 0.20 25	Библиотека Matplotlib и создание информативных графиков для анализа данных	2	Практикум	Кабинет информати ки №301	Решение задач
5	01.1 1.20 25	Машинное обучение: основы	2	Лекция с элементами дискуссии	Кабинет информати ки №301	Устный опрос
6	08.1 1.20 25	Машинное обучение: задачи классификации (часть 1)	2	Практикум	Кабинет информати ки №301	Решение задач
7	15.1 1.20 25	Машинное обучение: задачи классификации (часть 2)	2	Практикум	Кабинет информати ки №301	Решение задач

8	22.1 1.20 25	Машинное обучение: задачи кластеризации (часть 1)	2	Практикум	Кабинет информати ки №301	Решение задач
9	29.1 1.20 25	Машинное обучение: задачи кластеризации (часть 2)	2	Практикум	Кабинет информати ки №301	Решение задач
10	06.1 2.20 25	Машинное обучение: линейная регрессия (часть 1)	2	Практикум	Кабинет информати ки №301	Решение задач
11	13.1 2.20 25	Машинное обучение: линейная регрессия (часть 2)	2	Практикум	Кабинет информати ки №301	Решение задач
12	20.1 2.20 25	Другие методы машинного обучения: деревья решений (часть 1)	2	Практикум	Кабинет информати ки №301	Решение задач

13	27.1 2.20 25	Другие методы машинного обучения: деревья решений (часть 2)	2	Практикум	Кабинет информати ки №301	Решение задач
14	10.0 1.20 26	Отбор и ранжирование признаков	2	Семинар- практикум	Кабинет информати ки №301	Устный опрос
15	17.0 1.20 26	Функции потерь и метрики качества	2	Семинар- практикум	Кабинет информати ки №301	Устный опрос
16	24.0 1.20 26	Композиция моделей	2	Практикум	Кабинет информати ки №301	Решение задач
17	31.0 1.20 26	Градиентный бустинг (часть 1)	2	Лекция- практикум	Кабинет информати ки №301	Тестовая работа по теме
18	07.0 2.20 26	Градиентный бустинг (часть 2)	2	Практикум	Кабинет информати ки №301	Тестовая работа по теме

19	14.0 2.20 26	Работа с текстами: парсинг сайтов (часть 1)	2	Практикум	Кабинет информати ки №301	Решение задач
20	21.0 2.20 26	Работа с текстами: парсинг сайтов (часть 2)	2	Практикум	Кабинет информати ки №301	Решение задач
21	28.0 2.20 26	Работа с текстами: парсинг сайтов (часть 3)	2	Практикум	Кабинет информати ки №301	Решение задач
22	07.0 3.20 26	Работа с текстами: создание тг- ботов (часть 1)	2	Практикум	Кабинет информати ки №301	Решение задач
23	14.0 3.20 26	Работа с текстами: создание тг- ботов (часть 2)	2	Практикум	Кабинет информати ки №301	Решение задач

24	21.0 3.20 26	Работа с текстами: создание тг- ботов (часть 3)	2	Практикум	Кабинет информати ки №301	Решение задач
25	28.0 3.20 26	Введение в глубинное обучение	2	Лекция с демонстрац ией	Кабинет информати ки №301	Устный опрос
26	04.0 4.20 26	Полносвязные слои, сверточные слои (часть 1)	2	Лекция- практикум	Кабинет информат ики №301	Устный опрос
27	11.0 4.20 26	Полносвязные слои, сверточные слои (часть 2)	2	Практикум	Кабинет информат ики №301	Тестова я работа по теме
28	18.0 4.20 26	Подготовка к защите проектов. Консультации	2	Индивиду альная работа	Кабинет информат ики №301	-
29	25.0 4.20 26	Презентация собственных исследований	2	Защита проектов	Кабинет информат ики №301	Защита проекта

30	30.0	Резервный день.
	4.20	Дополнительная
	26	защита проектов

2.2. Условия реализации программы

- Материально-техническое обеспечение: занятия проводятся в компьютерном классе, оснащенном персональными компьютерами (ноутбуками) с доступом в сеть Интернет, проектором и экраном. Компьютеры должны иметь предустановленное программное обеспечение: ОС (Windows/Linux), Python 3.7+, Jupyter Notebook, необходимые библиотеки (NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, BeautifulSoup4, NLTK, TensorFlow/Keras или PyTorch).
- Информационное обеспечение: используются аудио- и видеоматериалы, онлайн-документация по библиотекам Python, специализированные образовательные платформы (Stepik, Coursera), научные статьи и обзоры по тематике машинного обучения.
- Кадровое обеспечение: программу реализует педагог дополнительного образования, имеющий высшее образование в области информатики, математики или смежных областях, владеющий языком Python и имеющий практический опыт в области машинного обучения и анализа данных.

2.3. Формы аттестации

Для контроля достижения планируемых результатов используются следующие формы аттестации:

- Текущая аттестация: устный опрос, решение практических задач, тестирование.
- Промежуточная аттестация: тестовые работы по ключевым модулям (например, по теме "Градиентный бустинг").
- Итоговая аттестация: защита индивидуального или группового проекта.

2.4. Оценочные материалы

Оценочные материалы включают в себя:

- Банк практических заданий по каждой теме.
- Тестовые задания для проверки теоретических знаний.
- Критерии оценки итогового проекта (актуальность, сложность, качество кода, оформление, результаты, презентация).

2.5. Методические материалы

- Конспекты лекций и презентации по всем темам курса.
- Практические работы и руководства по их выполнению.
- Примеры кода и шаблоны для проектов.
- Рекомендации по использованию современных педагогических технологий: проблемное обучение, проектная деятельность, дифференцированный подход.

2.6. Раздел о воспитании

Реализация программы направлена на достижение следующих воспитательных результатов:

- Гражданско-патриотическое воспитание: обсуждение роли российских ученых и IT-компаний в развитии мировых технологий ИИ, осознание важности technological sovereignty (технологического суверенитета).
- Духовно-нравственное воспитание: формирование этического отношения к использованию ИИ и данных, обсуждение проблем bias (смещения) в алгоритмах и ответственности разработчика.
- Воспитание культуры здорового и безопасного образа жизни: соблюдение правил работы за компьютером, профилактика цифрового переутомления.
- Трудовое воспитание: формирование ответственности, дисциплинированности, трудолюбия и целеустремленности при работе над сложными и долгосрочными проектами.
- Эстетическое воспитание: развитие чувства гармонии и ясности при визуализации данных и оформлении результатов проектной деятельности.

2.7. Список литературы

Для педагога:

1. Орельян, А. Глубокое обучение на Python. – СПб.: Питер, 2021.
2. Джеррард, П. Машинное обучение с использованием Python. – М.: ДМК Пресс, 2020.
3. Мюллер, А., Гвидо, С. Введение в машинное обучение с помощью Python. – М.: Вильямс, 2019.

Для учащихся:

1. Рассказов, П.А. Python и машинное обучение. – М.: Эксмо, 2022.
2. Чакрабарти, С. Анализ текста и машинное обучение на Python. – М.: ДМК Пресс, 2021.
3. Онлайн-документация по библиотекам NumPy, Pandas, Scikit-learn.

Для родителей:

1. Информационные бюллетени о целях и содержании программы, ее роли в профессиональной ориентации учащихся.

