

ПРОГРАММА
элективного курса

«Основы технологий искусственного интеллекта»

10-11 классы

64 часа

Москва, 2021

1. Планируемые результаты освоения учебного курса

Личностные:

- готовность и способность к самостоятельной, творческой и ответственной деятельности; навыки сотрудничества со сверстниками, детьми младшего возраста, взрослыми в образовательной, общественно полезной, учебно-исследовательской, проектной и других видах деятельности;
- готовность и способность к образованию, в том числе самообразованию, на протяжении всей жизни;
- сознательное отношение к непрерывному образованию как условию успешной профессиональной и общественной деятельности;
- осознанный выбор будущей профессии и возможностей реализации собственных жизненных планов; отношение к профессиональной деятельности как возможности участия в решении личных, общественных, государственных, общенациональных проблем.

Метапредметные:

- умение самостоятельно определять цели и составлять планы в различных сферах деятельности, осознавая приоритетные и второстепенные задачи; самостоятельно осуществлять, контролировать и корректировать учебную, внеурочную и внешкольную деятельность с учётом предварительного планирования;
- использовать различные ресурсы для достижения целей;
- выбирать успешные стратегии в трудных ситуациях;
- умение продуктивно общаться и взаимодействовать с коллегами по совместной деятельности, учитывать позиции другого (совместное целеполагание и планирование общих способов работы на основе прогнозирования, контроль и коррекция хода и результатов совместной деятельности), эффективно разрешать конфликты;

- владение навыками исследовательской и проектной деятельности (определение целей и задач, планирование проведения исследования, формулирование гипотез и плана их проверки);
- осуществление наблюдений и экспериментов, использование количественных и качественных методов обработки и анализа полученных данных; построение доказательств в отношении выдвинутых гипотез и формулирование выводов; представление результатов исследования в заданном формате, составление текста отчёта и презентации с использованием информационных и коммуникационных технологий).

Предметные:

- владение базовыми элементами теории вероятностей, методов математической статистики и методов машинного обучения;
- умение находить закономерности в данных, разрабатывать математические модели и модели машинного обучения на эти данных;
- умение выполнять численный анализ данных и визуализировать полученные результаты на языке Python;
- владение практическим опытом решения задач с применением методов математической статистики и машинного обучения.

2. Содержание учебного курса

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Основные принципы и подходы к моделированию. Классификация математических моделей. Бытовое понятие о вероятности. Сложение вероятностей совместных и несовместных событий. Перестановки, выборки и сочетания. Примеры. Понятие условной вероятности. Формула Байеса. Теорема о полной вероятности. Дискретная случайная величина. Схемы повторения испытаний. Формула Пуассона. Законы распределения дискретной случайной величины. Непрерывная случайная величина. Функция распределения. Плотность распределения. Равномерный закон распределения. Нормальный закон распределения. Что такое статистические оценки и чем занимается математическая статистика? Эмпирическая функция распределения. Принципы построения гистограмм. Ядерная оценка плотности распределения. Выборочное среднее. Выборочная дисперсия. Выборочное среднее квадратическое отклонение. Выбор функции распределения как вероятностной модели случайной величины. Вероятностная модель как смесь распределений. Смесь распределений Гаусса. Примеры. Точечные оценки параметров распределения. Метод максимального правдоподобия. Понятие вектора и матрицы. Операции над матрицами. Матричные произведения. Специальные виды матриц. Обратная матрица. Понятие СЛАУ. Методы решения СЛАУ (обзорно). Метод Гаусса. Выборочные коэффициенты корреляции. Корреляционная матрица. Уравнение прямой и задача регрессии. Множественная регрессия. Оценка качества регрессионной модели. Понятие вероятностного интервала. Примеры. Расстояние между объектами. Типы кластеров. Методы кластеризации (обзорно). Метод ближайшего соседа. Метод k-means. Кластеризация с помощью вероятностных моделей: разделение смеси Гауссовых распределений (дискриминантный анализ). Методы классификации. Логистическая регрессия. Использование градиента в задачах оптимизации и машинного обучения. Метод градиентного спуска. Стохастический градиентный спуск. Примеры.

АНАЛИЗ И ВИЗУАЛИЗАЦИЯ ДАННЫХ НА PYTHON

Какие бывают данные. Понятия числовых, категориальных данных. Способы представления информации. Основные задачи анализа данных: классификация, регрессия, кластеризация (повторение). Понятие функции и аргумента. Зависимость и независимость. Построение графика функции по табличным значениям. Понятие гистограммы как способа представления табличных данных, примеры (повторение). Понятие распределения (повторение) и способы визуализации различных распределений. Базовые типы данных в Python: численные, строковые, логические переменные. Циклы. Функции. Структуры данных в Python: списки, множества и словари - примеры создания и основные операции с ними. Понятие list comprehension. Пример реализации функции одной переменной. Импорт модулей и функций. Основные конструкции библиотеки numpy как библиотеки для высокопроизводительных вычислений. Векторизация вычислений. Создание массивов, одномерные и многомерные массивы. Вычисление основных статистических показателей матрицы с помощью numpy: минимум, максимум, среднее, argmax и др. Примеры. Основные конструкции библиотеки pandas. Чтение файлов и запись в файл. Понятие pandas.DataFrame и pandas.Series. Выгрузка данных по условию. Создание таблиц. Агрегация и слияние имеющихся данных. Выполнение сложных запросов к датасету. Библиотека matplotlib и визуализация данных. Построение графика функции и создание своего стиля для графика. Линейные и логарифмические шкалы, выбор масштаба представления данных. Гистограммы в matplotlib. Примеры построения нескольких независимых графиков в одном окне: метод subplots(). Сохранение графиков в виде изображения. Понятие корреляции (повторение). Ложные корреляции. Виды зависимостей данных друг от друга. Понятие кросс-корреляции, автокорреляции и свёртки. Понятие ранговых списков. Корреляция Пирсона и корреляция Спирмена. Вычисление попарных корреляций и корреляционных таблиц средствами numpy и pandas. Heatmap и графическое представление таблиц данных. Задача обучения с учителем. Обучение по прецедентам. Объекты и целевые переменные. Понятие функции ошибок. Тренировочная и тестовая выборка. Задачи классификации и регрессии – сходства и различия. Данные для

обучения в виде таблиц значений. Задачи обучения без учителя и data mining. Обзор: кластеризация, корреляционный анализ, понижение размерности. Алгоритмы кластеризации (повторение). Библиотека scikit-learn и её использование для кластеризации данных в Python. Изменение параметров методов кластеризации и проверка качества кластеризации. Метрики кластеризации и их реализация в Python. Реализация алгоритмов KNN, SVM и Kmeans в библиотеке scikit-learn. Примеры и визуализация. Понятие линейной регрессии (повторение). Понятие весовых коэффициентов и настройка параметров модели. Отбор признаков и работа с данными. Скалирование и центрирование данных. недообучение и переобучение. Понятие регуляризации. Регуляризация модели линейной регрессии – подходы Lasso и Ridge и их отличия. Случай нелинейной зависимости, полиномиальная регрессия. Примеры и упражнения на Python. Логистическая регрессия как модель бинарной классификации. Целевая переменная и виды функции ошибок для задач классификации. Понятие функции активации и её виды: линейная, сигмоида, гиперболический тангенс, ReLU. Примеры и упражнения на Python. Изображение как матрица. Понятие RGB изображения и примеры других цветовых пространств. Понятие яркости и контраста. Основы обработки изображений: фильтрация, бинаризация, выделение границ, размытие. Загрузка изображений в Python и использование библиотеки matplotlib для работы с изображениями. Примеры и упражнения по обработке изображений в Python.

ПАРАЛЛЕЛЬНАЯ ОБРАБОТКА И УПРАВЛЕНИЕ БОЛЬШИМИ ДАННЫМИ

Обработка цифровой, символьной, текстовой и табличной информации. Реляционные и NoSQL базы данных, их отличия, области применения, примеры использования. Строковые, целочисленные, дробные, дата, время. Понятие таблицы, ключа. Нормальные формы. Ключи, первичные и внешние ключи. Создание таблиц, вставка, выборка, удаление, изменение данных. Создание ключей на колонки. Выборка данных из таблиц, фильтрация, сортировки, группировки, слияние, подзапросы. Основные задачи и этапы проектирования баз данных. Концептуальное, Логическое, Физическое проектирование. Чтение данных, обработка и запись в различные форматы. Достоинства и недостатки распределенной обработки данных. Способы распределения

данных: централизованный, децентрализованный, смешанный. Компоненты экосистемы Apache Spark, Особенности Apache Spark, RDD и особенности использования, трансформации и действия. Знакомство со Spark-shell, Написание программ в Apache Spark, чтение и запись данных. Понятие DataFrame, Использование DataFrame вместо RDD, простые запросы, фильтрация и агрегация. Продвинутые операции: join, broadcast, udf, udaf.

ВВЕДЕНИЕ В МАШИННОЕ ОБУЧЕНИЕ

Классификация моделей искусственного интеллекта по Расселу и Норвигу. Имитация когнитивных функций человека современными моделями машинного обучения. Определения машинного обучения. Опыт, задача, качество решения. Способы задания входных данных для алгоритма машинного обучения. Обобщающая способность модели. Дилемма смещения-разброса, понятие недообучения и переобучения. Обучение с учителем (карта методов). Разметка данных. Функции потерь. Тренировочная, тестовая (контрольная), валидационная (проверочная) выборка. Кросс-валидация. Метрики качества бинарной классификации. ROC-AUC и Precision-Recall кривые. Метрики качества для несбалансированных выборок. Обучение без учителя. Метрики качества для оценки результатов кластеризации. Модулярность. Коэффициент силуэта. Semi-supervised обучение. Обучение с подкреплением. Метрики расстояния (Манхэттенская, евклидово расстояние, косинусное расстояние). Метод ближайших соседей. Подбор числа соседей. Метод опорных векторов для случая линейно разделимой выборки. Решающие правила. Конструирование решающих правил. Решающие деревья. Метрики информативности. Подрезка решающих деревьев. Бэггинг. Случайный лес. Бустинг. Алгоритм AdaBoost. Обучение смеси гауссианов. EM-алгоритм. Алгоритмы тематического моделирования. Вероятностный латентно-семантический анализ. Алгоритм K средних. Применение EM-алгоритма для алгоритма K средних. Иерархическая кластеризация. Интерпретация дендрограмм. Кластеризация на графах. Методы детектирования аномалий. Многомерное шкалирование. Метод главных компонент. Методы обучения представлений для текстовых данных. Методы обучения представлений для графовых

данных. Агент и среда. Система подкрепления. Способы обучения с подкреплением. Задача о многоруком бандите. Q-обучение. Типология нейронных сетей. Однослойные модели нейронных сетей. Правило Хебба. Карты Кохонена. Алгоритм обратного распространения ошибки. Способы борьбы с переобучением для нейронных сетей. Принцип построения иерархических признаков. Архитектура сверточной нейронной сети. Слои свертки и субдискретизации. Реализация операций пулинга. Современные архитектуры сверточных нейросетей: ImageNet, VGG16. Архитектура сети RNN. Архитектура сети LSTM. Архитектура сети GRU. Примеры использования рекуррентных сетей в области машинного перевода и прогнозирования временных рядов. Ограниченная машина Больцмана. Автоэнкодеры. Сети глубокого доверия. Генеративно-сопоставительная сеть.

3. Поурочное планирование

№ п/п	Темы	Количество часов
I. Математические основы технологий искусственного интеллекта		
1.	Вводное занятие. Что такое математическая модель?	1
2.	Исчисление вероятностей и элементы комбинаторики	1
3.	Понятие случайной величины. Условная и полная вероятность	1
4.	Обработка результатов наблюдений. Понятие статистической оценки	1
5.	Числовые оценки выборочных характеристик	1
6.	Вероятностные модели случайной величины	1
7.	Оценка параметров распределения случайной величины	1
8.	Базовые понятия из линейной алгебры	1
9.	Элементы многомерного статистического анализа и моделирования. Базовые элементы корреляционного анализа и регрессионного анализа	1
10.	Понятие классификации и кластеризации. Как связаны эти две задачи? Чем классификация отличается от регрессии?	1
11.	Понятие градиента	1
II. Анализ и визуализация данных на Python		
12.	Анализ данных. Примеры и задачи	1
13.	Одномерный анализ данных	1
14.	График функции. Гистограммы. Распределения	1
15.	Введение в Python. Базовые операции	1
16.	Библиотека numpy. Примеры	1
17.	Библиотека pandas. Примеры	1
18.	Библиотека matplotlib. Примеры	1
19.	Понятие корреляции. Примеры на pandas и numpy	2
20.	Обучение с учителем. Примеры	2

21.	Обучение без учителя. Примеры	2
22.	Кластеризация данных на Python	2
23.	Линейная регрессия на Python	2
24.	Логистическая регрессия на Python	2
25.	Работа с изображениями в Python	2
III. Параллельная обработка и управление большими данными		
26.	Понятие обработки данных. Виды обработки данных. Виды баз данных	1
27.	Типы данных, таблицы и отношения между ними. Реляционная модель данных	1
28.	Введение в SQL. Примеры в PostgreSQL	1
29.	Проектирование баз данных. Цели проектирования. Нормализация данных. Проектирование базы данных в PostgreSQL	1
30.	Обработка данных в памяти. Продвинутый pandas	1
31.	Основные понятия распределенной обработки данных	1
32.	Знакомство с Apache Spark (PySpark)	1
33.	Параллельная и распределенная обработка больших данных средствами PySpark	1
34.	Проектирование хранилища и процесса обработки данных	1
IV. Введение в машинное обучение		
35.	Введение в машинное обучение	1
36.	Типология и метрики качества алгоритмов машинного обучения	2
37.	Метрические алгоритмы классификации	2
38.	Логические алгоритмы классификации Введение в ансамблевые методы	1
39.	Модели смесей распределений	1
40.	Методы кластеризации и детектирования аномалий	2
41.	Методы снижения размерности многомерных данных	2
42.	Обучение с подкреплением	2
43.	Введение в нейронные сети	2
44.	Многослойный перцептрон	2

45.	Свёрточные нейронные сети	2
46.	Рекуррентные нейронные сети	2
47.	Глубокое обучение без учителя	2
Итого часов:		64