

ПРИНЯТА
Педагогическим советом АНО ОШ ЦПМ
(протокол от 28 августа 2023 г. № 73)

УТВЕРЖДЕНА
приказом директора АНО ОШ ЦПМ
от 29 августа 2023 г. № 408

РАБОЧАЯ ПРОГРАММА
курса внеурочной деятельности «Анализ данных»
для обучающихся 10 класса

Москва, 2023 год

Пояснительная записка

Данный курс внеурочной деятельности рассчитан на изучение нового материала по теме «Анализ данных», который базируется на классическом курсе по теории вероятности и статистике, и на получение практических навыков применения теоретических знаний с использованием языка программирования Python. Целью данного курса не является строгое изучение математической статистики с формальными математическими доказательствами всех утверждений, поэтому для прохождения курса не требуются знания математического анализа, все необходимые математические инструменты будут даны слушателям в процессе освоения курса. Статистические методы обработки данных будут представлены исключительно с прикладной точки зрения. Успешное прохождение курса предполагает умение выбирать необходимые статистические инструменты в зависимости от поставленной задачи и применение этих инструментов с использованием языка программирования Python. Курс включает в себя также изучение стандартных инструментов и библиотек языка программирования Python, поэтому знание синтаксиса языка не является необходимым для прохождения курса. Курс знакомит с элементами современного машинного обучения, однако не является самостоятельным курсом по машинному обучению, а скорее базой для его дальнейшего изучения. Курс не предполагает решение задач по олимпиадному программированию. В заключительной части курса все его слушатели выполняют групповые проекты по анализу данных. Оценка за курс вычисляется по формуле , где – средняя оценка за домашние задания, – средняя оценка за контрольные работы, – итоговая оценка за финальный групповой проект.

Методологической основой реализации поставленной цели являются следующие принципы:

– Принцип развития, который состоит в том, что олимпиадная подготовка должна быть нацелена прежде всего на создание условий для всестороннего развития мышления и личностных качеств каждого ученика, а не ограничиваться тренингом в освоении ими методов олимпиадной математики. Суть этого принципа можно кратко выразить тезисом: «развитие средствами олимпиадной математики каждого ученика».

– Принцип «выращивания» состоит в совмещении, с одной стороны, внутренней активности ученика, его целенаправленных попыток раскрыть и реализовать свой потенциал, а с другой стороны, внешней организации этой активности со стороны учителя в рамках той же цели.

– Принцип успешности состоит в акцентировке на успешность, то есть в создании такой среды, где к ошибке относятся как к ступеньке роста, а не поводу для огорчения и порицания, где ценится и поддерживается успех каждого ученика относительно себя, независимо от начального уровня его подготовки и математических способностей.

Планируемые результаты освоения курса

Освоение курса внеурочной деятельности «Анализ данных» обеспечивает достижение следующих предметных образовательных результатов:

- знание классических теорем, фактов и методов из теории вероятности и статистики;
- умение применять статистические инструменты для математических задач и обосновывать допустимость их использования;
- умение использовать необходимый математический аппарат;

- умение видеть возможность применения тех или иных статистических инструментов для ситуаций из «жизни»;
- понимание практической значимости статистических инструментов и теорем;
- умение интерпретировать статистические показатели;
- умение проводить статистические эксперименты и интерпретировать результаты;
- умение выдвигать и проверять гипотезы относительно собранных данных;
- умение критично смотреть на данные: замечать закономерности, но при этом аккуратно экстраполировать результаты наблюдений, замечать несоответствия и противоречия;
- умение решать исследовательские задачи, результат которых заранее неизвестен;
- умение подготавливать данные для их дальнейшей обработки;
- повышение навыков работы с языком Python;
- умение обрабатывать и визуализировать данные с помощью языка Python;
- навыки работы в команде: планирование, распределение обязанностей, обсуждение и соотнесение полученных результатов друг с другом;
- навык самостоятельной проектной работы: постановка задачи, выполнение поставленной задачи с нуля, интерпретация результатов и презентация своего проекта.

Содержание курса

Раздел 1. Данные и их первичный анализ.

Что такое данные? Основные принципы сбора данных. Наблюдение vs эксперимент. Простейшие статистические характеристики данных (среднее, медиана, мода, среднеквадратичное отклонение).

Раздел 2. Классическая теория вероятностей.

Определение вероятностного пространства. Классическая вероятность. Геометрическая вероятность. Байесовский подход к вероятности. Основные теоремы на подсчет вероятностей. Произведение и сумма событий. Построение отрицаний. Вероятности дополнительных событий. Независимые события. Вероятность произведения событий. Вероятность суммы событий. Условная вероятность. Формула полной вероятности. Теорема Байеса. Парадокс теоремы Байеса. Априорные и апостериорные вероятности. Наивный классификатор Байеса. ROC/AUC кривые.

Раздел 3. Распределения и основные статистические теоремы.

Случайная величина. Распределение. Дискретные распределения. Основные характеристики распределения: математическое ожидание, дисперсия. Подсчет характеристик распределения. Плотность распределения. Непрерывные распределения. Нормальное распределение. Z-оценка. Центральная предельная теорема. Бернулли и геометрическое распределение. Биномиальное распределение. Нормальная аппроксимация биномиального распределения. Пропорциональность распределения случайного образца.

Раздел 4. Тестирование гипотез и статистических предположений.

Оценка неизвестного параметра. Несмещенная оценка. Метод максимального правдоподобия. Интервалы доверия. Тестирование статистических гипотез. Нулевая

и альтернативная гипотезы. Оценка гипотезы и статистическая значимость. Ошибки первого и второго рода. Интервал доверия для пропорции. Тестирование гипотез для пропорции. Интервал доверия для разницы двух пропорций. Хи-квадрат распределение. Критерий хи-квадрат. Проверка гипотезы однородности. Проверка гипотезы независимости. Распределение Стьюдента. Оценка параметров с использованием распределения Стьюдента. Многофакторный дисперсионный анализ.

Раздел 5. Регрессии.

Корреляция. Линейная аппроксимация. Метод наименьших квадратов. Категориальный предиктор. Линейная регрессия. Логистическая регрессия. Цепи Маркова.

Раздел 6. Групповой проект.

Подготовка теоретической базы проекта. Сбор данных для проекта. Подготовка данных для анализа. Анализ собранных данных для проекта. Описание выводов проведенного анализа. Подготовка итоговых презентаций к защите проекта. Защиты проектов.

Тематическое планирование курса

| № | Наименование раздела / темы | Количество ак. ч. | Ссылка на ЭОР |
|--|--|-------------------|---|
| 1 | Данные и их первичный анализ | 8 | |
| 1.1 | Что такое данные? Основные принципы сбора данных. Наблюдение vs эксперимент. | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 1.2 | Простейшие статистические характеристики данных (среднее, медиана, мода, среднеквадратичное отклонение). | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 2 | Классическая теория вероятностей | 12 | |
| 2.1 | Определение вероятностного пространства. Классическая вероятность. Геометрическая вероятность. Байесовский подход к вероятности. Основные теоремы на подсчет вероятностей. Произведение и сумма событий. Построение отрицаний. Вероятности дополнительных событий. Независимые события. Вероятность произведения событий. Вероятность суммы событий. | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 2.2 | Условная вероятность. Формула полной вероятности. Теорема Байеса. Парадокс теоремы Байеса. Априорные и апостериорные вероятности. Наивный классификатор Байеса. | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 2.3 | ROC/AUC кривые | 4 | http://school-collection.edu.ru/collection/matematika/ |
| Контрольная работа по разделам 1 – 2 | | 2 | |
| Разбор контрольной работы по разделам 1 – 2, работа над ошибками | | 2 | |

| | | | |
|---|---|-----------|---|
| 3 | Распределения и основные статистические теоремы | 24 | |
| 3.1 | Пределы | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 3.2 | Производные | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 3.3 | Интегралы | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 3.4 | Случайная величина. Распределение. Дискретные распределения. Основные характеристики распределения: математическое ожидание, дисперсия. Подсчет характеристик распределения. | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 3.5 | Плотность распределения. Непрерывные распределения. | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 3.6 | Нормальное распределение. Z-оценка. Центральная предельная теорема. Бернулли и геометрическое распределение. Биномиальное распределение. Нормальная аппроксимация биномиального распределения. Пропорциональность распределения случайного образца. | 4 | http://school-collection.edu.ru/collection/matematika/ |
| Контрольная работа по разделу 3 | | 2 | |
| Разбор контрольной работы по разделу 3, работа над ошибками | | 2 | |
| 4 | Тестирование гипотез и статистических предположений | 20 | |
| 4.1 | Оценка неизвестного параметра. Несмещенная оценка. Метод максимального правдоподобия. Интервалы доверия. | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 4.2 | Тестирование статистических гипотез. Нулевая и альтернативная гипотезы. Оценка гипотезы и статистическая значимость. Ошибки первого и второго рода. | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 4.3 | Интервал доверия для пропорции. Тестирование гипотез для пропорции. Интервал доверия для разницы двух пропорций. | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 4.4 | Хи-квадрат распределение. Критерий хи-квадрат. Проверка гипотезы однородности. Проверка гипотезы независимости. Распределение Стьюдента. Оценка параметров с использованием распределения Стьюдента. | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 4.5 | Многофакторный дисперсионный анализ | 4 | http://school-collection.edu.ru/collection/matematika/ |
| Контрольная работа по разделу 4 | | 2 | |

| | | | |
|---|---|------------|---|
| Разбор контрольной работы по разделу 4, работа над ошибками | | 2 | |
| 5 | Регрессии | 16 | |
| 5.1 | Корреляция. Линейная аппроксимация. Метод наименьших квадратов. | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 5.2 | Категориальный предиктор. Линейная регрессия. | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 5.3 | Логистическая регрессия | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 5.4 | Цепи Маркова | 4 | http://school-collection.edu.ru/collection/matematika/ |
| Контрольная работа по разделу 5 | | 2 | |
| Разбор контрольной работы по разделу 5, работа над ошибками | | 2 | |
| 6 | Групповой проект | 40 | |
| 6.1 | Подготовка теоретической базы проекта | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 6.2 | Сбор данных для проекта | 4 | http://school-collection.edu.ru/collection/matematika/ |
| 6.3 | Подготовка данных для анализа | 6 | http://school-collection.edu.ru/collection/matematika/ |
| 6.4 | Анализ собранных данных для проекта | 6 | http://school-collection.edu.ru/collection/matematika/ |
| 6.5 | Описание выводов проведенного анализа | 6 | |
| 6.6 | Подготовка итоговых презентаций к защите проекта | 6 | |
| 6.7 | Защиты проектов | 8 | |
| ВСЕГО | | 136 | |

Учебно-методическое обеспечение образовательного процесса

Цифровые образовательные ресурсы сети Интернет

1. Информационная система "Единое окно доступа к образовательным ресурсам" - <http://window.edu.ru/>;

2. Единая коллекция цифровых образовательных ресурсов - <http://school-collection.edu.ru/>;

3. Федеральный центр информационно-образовательных ресурсов - <http://fcior.edu.ru/>.

4. Электронный архив научно – популярных журналов <http://library.controlchaostech.com/bel/>

5. Список сайтов, рекомендованных для использования обучающимися и преподавателями для доступа к высококачественным ЭОР

<http://www.mon.gov.ru/> - Министерство образования и науки Российской Федерации

<http://www.uznai-prezidenta.ru/> - Детский сайт Президента Российской Федерации

<http://www.eidos.ru/olymp/> - Всероссийские дистанционные эвристические олимпиады

<http://www.en.edu.ru> - Естественнонаучный образовательный портал

<http://ege.edu.ru>- Портал информационной поддержки Единого государственного экзамена

<http://www.school.edu.ru> - Российский общеобразовательный портал

<http://vschool.km.ru> - Виртуальная школа Кирилла и Мефодия

www.school-collection.ru - Единая коллекция цифровых образовательных ресурсов

<http://www.edu.ru> - Федеральный портал «Российское образование»

<http://school-collection.edu.ru/collection/matematika/> – материалы по математике в Единой коллекции цифровых образовательных ресурсов

<http://www.maht-on-line.com> – Занимательная математика – школьникам (олимпиады, игры, конкурсы по математике)

<http://www.mathkang.ru> – международный математический конкурс «Кенгуру»

<http://http://ege2011.mioo.ru> – Московский институт открытого образования, система СтатГрад