

ПРИНЯТА

Педагогическим советом АНО ОШ ЦПМ
(протокол от 28 августа 2024 г. №99)

УТВЕРЖДЕНА

приказом директора АНО ОШ ЦПМ
от 29 августа 2024 г. №677-ОД24

РАБОЧАЯ ПРОГРАММА

курса внеурочной деятельности «Введение в анализ данных»
для обучающихся 10 класса

Пояснительная записка

Программа курса «Введение в анализ данных» составлена на основе положений и требований к результатам освоения основной образовательной программы, представленных в ФГОС СОО и с учётом федеральной программы воспитания.

Введение в анализ данных – это курс, позволяющий школьнику выйти за пределы школьной программы и понять, как функционирует современная наука, рынок труда и применять знания, полученные на других курсах (математика, обществознание, количественные методы исследований). Курс позволяет привить любовь к обозначенным выше дисциплинам, осознать важность их изучения и стать более конкурентноспособным в интеллектуальных соревнованиях школьников. С одной стороны, динамичное развитие современного мира вынуждают преподавателей регулярно находить новые материалы, чтобы поддерживать интерес обучающихся к предмету и успевать за трендами, меняющими наш мир, с другой – интеллектуальные соревнования школьников также регулярно создают задания, для решений которых важно знать более сложный материал. Курс позволяет решить обе проблемы: и оживление учебной программы для рассказа о новых трендах в образовании, и подготовка учеников к интеллектуальным соревнованиям (олимпиадам и конкурсам). Более того, он позволяет школьникам расширить спектр олимпиад, доступных им для участия (например, олимпиада DANO).

Учитывая стремительное повышение спроса на специалистов в сфере анализа данных, программирования и работы с данными, образование старается интегрировать в образовательный процесс новые предметы, дающие соответствующие компетенции. Запускаются федеральные проекты, поддерживающие развитие IT-индустрии, а также крупные игроки банковского сектора инвестируют в развитие сферы анализа данных и программирования. Ученики, прошедшие курс «Введение в анализ данных», освоят базовую работу с кодом, научатся продвинутым методам анализа данных, а также разберутся в основах машинного обучения. Перечисленные выше навыки крайне важны для современного экономиста, который в будущем будет сталкиваться с вызовами, от преодоления которых вполне возможно будет зависеть будущее нашей страны.

Целями курса является:

Углубление знаний о математической статистике, работа в python, работа в excel, работа в orange, изучение современных статистических инструментов анализа данных, погружение в методы машинного обучения;

Погружение учеников в новые технологии, рассказ о их роли в жизни общества и новых вызовах, которые появляются в связи с развитием технологий;

развитие способности обучающихся к личному самоопределению, самореализации, самоконтролю;

развитие интереса обучающихся к освоению социальных и гуманитарных дисциплин;

овладение умениями получать, анализировать, интерпретировать;

систематизировать социальную информацию из различных источников, преобразовывать ее и использовать для самостоятельного решения учебно-познавательных, исследовательских задач, а также в проектной деятельности.

Освоение содержания курса «Введение в анализ данных» осуществляется в

соответствии со следующими ориентирами, отражающими специфику учебного предмета на уровне среднего общего образования:

определение учебного содержания научной и практической значимостью включаемых в него положений и педагогическими целями учебного предмета с учетом познавательных возможностей учащихся старшего подросткового возраста;

представление в содержании учебного предмета основных сфер жизни общества, типичных видов человеческой деятельности в информационном обществе, перспектив и прогнозов общественного развития, путей решения актуальных социальных проблем;

обеспечение развития ключевых навыков, формируемых деятельностным компонентом социально-математического образования (выявление проблем, принятие решений, работа с информацией, использование знаний математики в анализе данных), и компетентностей, имеющих универсальное значение для различных видов деятельности и при выборе профессии;

расширение возможностей самопрезентации обучающихся, мотивирующей креативное мышление и участие в социальных практиках.

Отличие содержания курса на базовом уровне среднего общего образования от содержания альтернативных курсов заключается в:

изучении нового теоретического содержания;

изучении новых сфер знания (статистика, анализ данных, введение в программирование);

освоении обучающимися базовых методов социального и аналитического познания;

большей опоре на самостоятельную деятельность и индивидуальные познавательные интересы обучающихся, в том числе связанные с выбором профессии;

расширении и совершенствовании познавательных, исследовательских, проектных умений, которые осваивают обучающиеся, и возможностей их применения при выполнении социальных ролей, типичных для старшего подросткового возраста.

Содержание курса

Раздел I. Excel для анализа данных

I.1. Начало работы в Excel

Пакет Microsoft Office, работа с табличными данными, интерфейс Excel, ячейки, формулы.

I.2. Табличные данные в Excel

Форматы данных, очистка данных, условное форматирование, фильтры и сортировка.

I.3. Реализация вычислений в Excel

Очистка данных. Основные статистические формулы, логические функции, математические формулы. Интерпретация статистических показателей.

I.4. Сводные таблицы в Excel

Создание и редактирование сводных таблиц, расчёт показателей без формул с помощью сводных таблиц.

I.5. Визуализация в Excel

График распределения. Диаграмма, линейный график, барчарт vs. гистограмма, ящик с усами, диаграмма рассеяния. Основные принципы визуализации данных. Интерпретация графиков. Перенос и использование визуализации в Power Point.

Раздел II. Введение в Orange

II.1. Начало работы в Orange

Интерфейс, загрузка данных, функционал.

II.2. Визуализация данных в Orange

Диаграмма, линейный график, барчарт vs. гистограмма, ящик с усами, диаграмма рассеяния. Интерпретация графиков.

II.3. Анализ данных в Orange

Расчёт основных статистических показателей. Корреляция. Кластерный анализ. Интерпретация результатов.

Раздел III. Основы программирования в Python

III.1. Начало работы в Python

Среда программирования, переменные. Инпут. Интеджер, флот, строка. Математические операции.

III.2. Структурированные типы данных, логические переменные

Список, кортеж. Булев тип данных. Условные операторы. Конструкция if-else.

III.3. Циклы

Цикл с условием, логические выражения в условии цикла while, бесконечный цикл. Работа с последовательностью. Использование цикла for для просмотра списков и кортежей, перебора индексов и числовых промежутков, редактирования списков. Вложенный цикл.

III.4. Неупорядоченные типы данных

Множество, словарь. Операции над множествами, практическое применение множеств. Добавление элементов в словарь, поиск по словарю. Подсчёт частот с помощью словаря.

III.5. Работа с текстовыми файлами

Чтение и запись в файл. Методы анализа текстовых данных, количественный контент-анализ. Работа с csv-файлами. Модуль csv.

III.6. Функции и библиотеки

Разбиение программы на функции. Lambda-функции. Использование библиотек, основные библиотеки для анализа данных.

III.7. Pandas и очистка данных

Модуль Pandas. Индексация, фильтрация, сортировка. Редактирование и сохранение таблиц. Перекодирование переменных.

III.8. Реализация расчётов в Python

Очистка данных. Расчёт основных статистических показателей. Интерпретация показателей.

III.9. Визуализация данных в Python

Модуль matplotlib. Модуль seaborn. Редактирование и сохранение визуализаций. Интерпретация графиков.

Раздел IV. Основы машинного обучения

IV.1. Введение в машинное обучение и анализ данных

Разбор библиотек pandas и numpy. Практикум, решение задач обработки данных и первичного анализа. Разведывательный анализ с помощью обозначенных библиотек.

IV.2. Виды задач и виды признаков. Метод k ближайших соседей для классификации. Вычисление расстояний для числовых данных

Определение признаков, разбор примеров. Введение в метод ближайших соседей. Теоретическое погружение, математический смысл. Практика в python.

IV.3. Измерение ошибки в задачах классификации. Обобщающая способность. Проверка обобщающей способности: отложенная выборка и кросс-валидация.

Классификация: виды и особенности. Примеры использования в академическом исследовании и прикладных задачах. Работа с выборками. Практикум кросс-валидации выборки на реальных данных.

IV.4. Модель линейной регрессии.

Подготовка данных для линейной регрессии. MSE для линейной регрессии в матричной форме. Математический смысл регрессии и подсчет MSE.

IV.5. Модель линейной регрессии.

Подготовка данных для линейной регрессии. Переобучение и регуляризация линейных моделей. Практикум в python.

IV.6. Линейная классификация.

Математический разбор линейной классификации. Метрики качества классификации. Практикум в python.

IV.7. Алгоритмы работы с текстами.

Разбор основных алгоритмов работы с текстами. Важность и актуальность работы с текстами для аналитика данных. Изучение токенов. Практика в python.

IV.8. Решающие деревья.

Предсказание значений целевой переменной с помощью применения последовательности простых решающих правил. Жадный алгоритм построения решающего дерева. Критерии ветвления: общая идея. Информативность в задаче регрессии: MSE. Особенности данных. Практикум в python.

IV.9. Градиентный бустинг.

Математическое обоснование. Обучение базового алгоритма. Практика и темп обучения. Градиентный бустинг и нейросети.

Планируемые результаты

В результате прохождения курса «Введение в анализ данных» у учеников появится понимание современной науки и рынка труда: ученики будут иметь возможность выйти за пределы школьной программы и понять, как функционирует современная наука, рынок труда

и применять знания, полученные на других курсах (математики, обществознания).

Помимо этого, произойдет развитие любви к математике и обществознанию: курс позволяет привить любовь к обозначенным выше дисциплинам, осознать важность их изучения. Важным результатом курса будет считаться подготовка к интеллектуальным соревнованиям: ученики станут более конкурентноспособными в интеллектуальных соревнованиях, конкурсах и олимпиадах. Новые олимпиады, набирающие популярность и связанные с анализом данных, требуют знаний в python, уверенной работе в excel и в целом понимание того, как необходимо работать с данными.

Отдельно важно отметить ознакомление с новыми трендами в образовании: динамичное развитие современного мира вынуждают преподавателей регулярно находить новые материалы, чтобы поддерживать интерес обучающихся к предмету и успевать за трендами, меняющими наш мир. Следующий результат – это приобретение знаний о количественных методах исследования: ученики узнают, что такое доверительный интервал, регрессионный анализ, современные методы количественного исследования, а также базовые категории математической статистики.

Ученики научатся использовать python для анализа данных, реализации собственных проектов и презентации результатов. Прикладная компонента курса поможет на практике отработать изученные материалы, а также подготовит учеников к более серьезному изучению машинного обучения. Полученные навыки и компетенции можно использовать как при сдаче индивидуального проекта, который входит в обязательный план обучающихся, так и при участии в интеллектуальных соревнованиях. Ученики в конце курса будут обладать навыками машинного обучения, работе в python, excel и latex.

Также важным результатом является развитие критического мышления: курс поможет ученикам развить критическое мышление и умение анализировать данные.

Дополнительным результатом можно считать улучшение учебной успеваемости: изучение количественных методов исследования поможет ученикам лучше понимать математические и общественные науки, что в свою очередь может привести к улучшению учебной успеваемости. Плюс, подготовка к будущей профессиональной деятельности: изучение количественных методов исследования может быть полезно для будущей профессиональной деятельности учеников в таких областях, как экономика, социология, психология и другие.

Тематическое планирование курса

Раздел\тема	Количество ак. часов
Раздел I. Excel для анализа данных	28
I.1. Начало работы в Excel	2
I.2. Табличные данные в Excel	2
I.3. Реализация вычислений в Excel	6
I.4. Сводные таблицы в Excel	2
I.5. Визуализация в Excel	6
I.6. Пакет анализ данных в Excel	6

Защита итоговых проектов	4
Раздел II. Введение в Orange	16
2.1. Начало работы в Orange	2
2.2. Визуализация данных в Orange	4
2.3. Анализ данных в Orange	4
Контрольная работа	2
2.4. LaTeX	4
Раздел III. Основы программирования в Python	38
3.1. Начало работы в Python	2
3.2. Структурированные типы данных, логические переменные	4
3.3. Циклы	4
3.4. Неупорядоченные типы данных	4
3.5. Работа с текстовыми файлами	4
3.6. Функции и библиотеки	4
3.7. Pandas и очистка данных	4
3.8. Реализация расчётов в Python	4
3.9. Визуализация данных в Python	4
Защита итоговых проектов	4
Раздел IV. Основы машинного обучения	54
4.1. Введение в машинное обучение и анализ данных	4
4.2. Виды задач и виды признаков. Метод k ближайших соседей для классификации. Вычисление расстояний для числовых данных	6
4.3. Измерение ошибки в задачах классификации. Обобщающая способность. Проверка обобщающей способности: отложенная выборка и кросс-валидация.	8
4.4. Модель линейной регрессии.	6
4.5. Модель линейной регрессии.	4
4.6. Линейная классификация.	6
4.7. Алгоритмы работы с текстами.	8
4.8. Решающие деревья.	4
4.9. Градиентный бустинг.	6
Контрольная работа	2
ИТОГО	136