|  |  |
| --- | --- |
| **Национальный** **исследовательский университет** **«Высшая школа экономики»****Лицей** | **Приложение 605**УТВЕРЖДЕНОпедагогическим советом Лицея НИУ ВШЭпротокол № 10 от 26.04.2023 |

Рабочая программа по учебному предмету (курсу)

«Основы анализа данных на Python.

Направление «Информатика, инженерия и математика»

10-11 класс

Автор:

Палеев Д.А.

1. **Планируемые результаты освоения учебного предмета**

К **личностным результатам**, на становление которых оказывает влияние изучение курса «Основы анализа данных», можно отнести:

* сознательное самоопределение ученика относительно профиля дальнейшего обучения или профессиональной деятельности;
* принятие и реализация ценностей здорового и безопасного образа жизни, бережное, ответственное и компетентное отношение к собственному физическому и психологическому здоровью;
* готовность обучающихся к конструктивному участию в принятии решений, затрагивающих их права и интересы, в том числе в различных формах общественной самоорганизации, самоуправления, общественно значимой деятельности;
* нравственное сознание и поведение на основе усвоения общечеловеческих ценностей, толерантного сознания и поведения в поликультурном мире, готовности и способности вести диалог с другими людьми, достигать в нем взаимопонимания, находить общие цели и сотрудничать для их достижения;
* развитие компетенций сотрудничества со сверстниками, детьми младшего возраста, взрослыми в образовательной, общественно полезной, учебно-исследовательской, проектной и других видах деятельности.
* мировоззрение, соответствующее современному уровню развития науки, значимости науки, готовность к научно-техническому творчеству, владение достоверной информацией о передовых достижениях и открытиях мировой и отечественной науки, заинтересованность в научных знаниях об устройстве мира и общества;
* 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;
* уважение ко всем формам собственности, готовность к защите своей собственности;
* осознанный выбор будущей профессии как путь и способ реализации собственных жизненных планов;
* готовность обучающихся к трудовой профессиональной деятельности как к возможности участия в решении личных, общественных, государственных, общенациональных проблем.

**Метапредметные результаты** освоения образовательной программы по курсу «Основы анализа данных» отражают овладение универсальными учебными действиями — познавательными, коммуникативными, регулятивными.

*Универсальные познавательные действия*

* искать и находить обобщенные способы решения задач, в том числе, осуществлять развернутый информационный поиск и ставить на его основе новые (учебные и познавательные) задачи;
* критически оценивать и интерпретировать информацию с разных позиций, распознавать и фиксировать противоречия в информационных источниках;
* использовать различные модельно-схематические средства для представления существенных связей и отношений, а также противоречий, выявленных в информационных источниках;
* находить и приводить критические аргументы в отношении действий и суждений другого; спокойно и разумно относиться к критическим замечаниям в отношении собственного суждения, рассматривать их как ресурс собственного развития;
* выходить за рамки учебного предмета и осуществлять целенаправленный поиск возможностей для широкого переноса средств и способов действия.

*Универсальные коммуникативные действия*

* осуществлять деловую коммуникацию как со сверстниками, так и со взрослыми (как внутри образовательной организации, так и за ее пределами), подбирать партнеров для деловой коммуникации исходя из соображений результативности взаимодействия, а не личных симпатий;
* координировать и выполнять работу в условиях реального, виртуального и комбинированного взаимодействия;
* развернуто, логично и точно излагать свою точку зрения с использованием адекватных (устных и письменных) языковых средств.

*Универсальные регулятивные действия*

* самостоятельно определять цели, задавать параметры и критерии, по которым можно определить, что цель достигнута;
* оценивать возможные последствия достижения поставленной цели в деятельности, собственной жизни и жизни окружающих людей, основываясь на соображениях этики и морали;
* ставить и формулировать собственные задачи в образовательной деятельности и жизненных ситуациях;
* оценивать ресурсы, в том числе время и другие нематериальные ресурсы, необходимые для достижения поставленной цели;
* выбирать путь достижения цели, планировать решение поставленных задач, оптимизируя материальные и нематериальные затраты;
* организовывать эффективный поиск ресурсов, необходимых для достижения поставленной цели;
* сопоставлять полученный результат деятельности с поставленной заранее целью.

Планируемые **предметные результаты:**

* уметь использовать библиотеки Python для работы с данными: Numpy, Pandas ;
* уметь использовать библиотеки Python для визуализации данных: Seaborn, Matplotlib;
* уметь обрабатывать большие наборы данных;
* знать основные практики “Разведывательного анализа данных” ;
* знать основные понятия и постановки задач машинного обучения;
* уметь анализировать графики и делать выводы по ним;
* уметь воспроизводить принцип проверки статистических гипотез;
* формулировать задачу на языке статистической гипотезы;
* уметь производить проверку статистических гипотез с помощью корректно подобранных статистических критериев и интерпретировать полученные результаты;
* уметь строить линейную регрессионную модель и интерпретировать полученные результаты;
* уметь выбирать оптимальную модель машинного обучения для решения задачи и обосновывать её выбор;
* понимать принцип работы нейронных сетей;
* уметь работать с библиотекой обучения нейронных сетей Pytorch;
* уметь дообучать свёрточные нейронные сети под конкретные задачи;

Учебный курс «Основы анализа данных» ориентирован на учащихся, которые хотят развить практические навыки в сфере анализа данных и машинного обучения. Данная программа рассчитана на углубленное изучение учебного курса в течение 68 часов в 10 классе и 34 часов в 11 классе.

Учебный курс относится к циклу дисциплин по выбору.

Целью освоения курса «Основы анализа данных» является развитие практических навыков в сфере анализа данных и машинного обучения. В процессе изучения курса, обучающиеся познакомятся с современными библиотеками машинного обучения и обработки данных: Pandas, Numpy, Seaborn, SciPy и Pytorch, и научатся применять их для решения задач анализа данных, машинного и глубинного обучения. В рамках курса учащимся будут представлены примеры этапов решения задачи анализа данных таких как: предобработка данных, разведывательный анализ, выдвижение гипотез и их проверка, выбор и обучение модели машинного обучения. Также обучающиеся получат необходимые навыки для участия в олимпиадах по анализу данных и искусственному интеллекту.

1. **Содержание учебного предмета**

**10 класс.**

**Тема 1. Знакомство с Python.**

Базовые понятия языка: переменные, условия, циклы. Функции, методы и операторы.

Библиотеки Pandas, Numpy. Работа с датасетами, манипуляции с данными и получение базовых статистик.

Типы данных, создание новых признаков

**Тема 2. Основы теории вероятности и статистики**.

Генеральная совокупность и выборка: частотные таблицы и распределения

Матожидание и разброс

Z-оценка, корреляция, выбросы

Тестирование гипотез, непараметрический критерии

Линейная регрессия и логистическая регрессия

**Тема 3. Визуализация данных**

Работа с библиотеками Python для визуализации данных: Matplolip, Seaborn, Plotly.

Изучение видов графиков: столбчатая диаграмма, гистограмма, пузырьковая диаграмма, box plot, круговая диаграмма. Чтение графиков и получение выводов из них.

Визуализация гипотез и анализа данных для представления на публике.

**Тема 4. Основы математики машинного обучения.**

Линейная алгебра. Матрицы, векторы, скалярное произведение, линейная независимость, векторные пространства, определитель матрицы, обращение матрицы, решение системы линейных уравнений, метод наименьших квадратов, матричные разложения.

Математический анализ. Дифференциал и производная, дифференцирование функций, дифференцирование в частных производных, градиент функции.

**Тема 5. Машинное обучение.**

Знакомство с библиотекой SciPy.

Задачи классификации, регрессии, ранжирования, кластеризации.

Линейная регрессия, резуляризация, логистическая регрессия, метод опорных векторов, многоклассовая классификация, решающие дервья, градиентный бустинг, обучение ранжированию, рекомендательные системы.

**Тема 6. Основы глубинного обучения**

Автоматическое дифференцирование, полносвязные нейронные сети.

Оптимизаторы и принципы их работы. SGD, Adam, AdamW, Dropout, Нормализация

Архитектура сверхточных нейронных сетей.

Задачи компьютерного зрения: классификация, сегментация, детекция.

1. **Тематическое планирование**

**10 класс**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Тематическое содержание занятия** | **Обязательные часы** | **Основные виды деятельности** |
| 1 | Знакомство с Python | 6 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Решение задач. Выполнение практических заданий на компьютере. |
| 2 | Основы теории вероятности и статистики | 18 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Решение задач. Выполнение практических заданий на компьютере. |
| 3 | Основы математики машинного обучения | 20 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Решение задач. Выполнение практических заданий на компьютере. |
| 4 | Машинное обучение | 24 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Решение задач. Выполнение практических заданий на компьютере. |
|  | **Всего** | **68** |  |

**11 класс**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Тематическое содержание занятия** | **Обязательные часы** | **Основные виды деятельности** |
| 1 | Машинное обучение  | 10 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Решение задач. Выполнение практических заданий на компьютере. |
| 2 | Основы глубинного обучения | 22 | Лекция и дискуссия по обозначаемой преподавателем проблематике. Решение задач. Выполнение практических заданий на компьютере. |
| 5 | Резерв | 2 |  |
|  | **Всего** | **34** |  |

В воспитании, развитии и социализации обучающихся средствами курса «Основы анализа данных» приоритетом является создание благоприятных условий для развития ценности научного познания, гражданского, трудового и духовно-нравственного воспитания:

* интерес к обучению и познанию; любознательность; готовность и способность к самообразованию, исследовательской деятельности, осознанному выбору направленности и уровня обучения в дальнейшем;
* сформированность информационной культуры, в том числе навыков самостоятельной работы с учебными текстами, справочной литературой, разнообразными средствами информационных технологий, а также умения самостоятельно определять цели своего обучения, ставить и формулировать для себя новые задачи в учёбе и познавательной деятельности, развивать мотивы и интересы своей познавательной деятельности;
* интерес к практическому изучению профессий и труда в сферах профессиональной деятельности, связанных с информатикой, программированием и информационными технологиями, основанными на достижениях науки информатики и научно–технического прогресса;
* осознанный выбор и построение индивидуальной траектории образования и жизненных планов с учётом личных и общественных интересов и потребностей.
* представление о социальных нормах и правилах межличностных отношений в коллективе, в том числе в социальных сообществах; готовность к разнообразной совместной деятельности при выполнении учебных, познавательных задач, создании учебных проектов; стремление к взаимопониманию и взаимопомощи в процессе этой учебной деятельности; готовность оценивать своё поведение и поступки своих товарищей с позиции нравственных и правовых норм с учётом осознания последствий поступков.
* ориентация на моральные ценности и нормы в ситуациях нравственного выбора; готовность оценивать своё поведение и поступки, а также поведение и поступки других людей с позиции нравственных и правовых норм с учётом осознания последствий поступков; активное неприятие асоциальных поступков, в том числе в сети Интернет.

Выделение данного приоритета связано с особенностями обучающихся юношеского возраста: с их потребностью в жизненном самоопределении, в выборе дальнейшего жизненного пути, который открывается перед ними на пороге самостоятельной взрослой жизни.

**Дополнительные материалы**

**Учебно-методическое обеспечение образовательной деятельности**

1. Поляков К.Ю., Еремин, Е.А. Информатика. Углубленный уровень. Учебник для 11 класса. В 2-х частях. Часть 1. М.: БИНОМ Лаборатория знаний.

2. Поляков К.Ю., Еремин, Е.А. Информатика. Углубленный уровень. Учебник для 11 класса. В 2-х частях. Часть 2. М.: БИНОМ Лаборатория знаний.

3. Marc Peter Deisenroth. Mathematics for Machine Learning (2020). Cambridge University Press; 1st edition

4. Флах, П., Машинное обучение: наука и искусство построения алгоритмов, которые извлекают знания из данных, 2015

5. Шень А., Вероятность: Примеры и задачи. Издательство МЦНМО, 2016

6. Син Такахаси, Иноуэ Ирона. Регрессионный анализ, Додека XXI век, 2010.