

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ))



Программа итоговой (государственной итоговой) аттестации, как компонент образовательной программы высшего образования - программы бакалавриата по направлению подготовки 09.03.01 Информатика и вычислительная техника, утвержденной первым проректором РУТ (МИИТ) Тимониным В.С.

ПРОГРАММА ИТОГОВОЙ (ГОСУДАРСТВЕННОЙ ИТОГОВОЙ) АТТЕСТАЦИИ

ПОДГОТОВКА К СДАЧЕ И СДАЧА ГОСУДАРСТВЕННОГО ЭКЗАМЕНА

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Направленность (профиль): IT-сервисы и технологии обработки данных на транспорте

Квалификация выпускника: Бакалавр

Форма обучения: Очная

РУТ (МИИТ)

ID подписи: 35225

Подписал: директор Игольников Борис Владимирович

Дата: 02.10.2024

Программа итоговой (государственной итоговой) аттестации в виде электронного документа выгружена из единой корпоративной информационной системы управления университетом и соответствует оригиналу

1. Итоговая (государственная итоговая) аттестация по направлению подготовки 09.03.01 Информатика и вычислительная техника и направленности (профилю) IT-сервисы и технологии обработки данных на транспорте в соответствии с учебным планом проводится в форме: Государственного экзамена / итогового экзамена (далее — экзамен).

2. Экзамен.

2.1. Программа экзамена.

Программа государственного экзамена включает в себя задания по дисциплинам, освоение которых имеет определяющее значение для профессиональной компетенции выпускников. Программа экзамена:

1. Основы web-программирования

1.1 PHP. Передача параметров

1.2 PHP. Графика и работа с файлами

1.3 PHP. Работа с базой данных

1.4 HTML. Табличная верста

1.5 HTML. Создание форм

1.6 CSS. Форматирование текста

1.7 CSS. Блочная верстка

2. Системы программирования Python

2.1 Классы в Python

2.2 Основные библиотеки для анализа данных, настройка IDE

2.3 Подключение к базе данных

2.4 Программный и математический аппарат для анализа данных

3. Машинное обучение и анализ данных

3.1 EDA\Feature engineering

3.2 Кластеризация

3.3 Классификация

3.4 Регрессия

3.5 Интерпретация моделей машинного обучения

4. Анализ больших текстовых данных и текстовый поиск

4.1 Seq2Seq модели с вниманием

4.2 Трансформеры

4.3 LLM

5. Алгоритмы и структуры данных

5.1 Алгоритмы

5.2 Структуры данных

5.3 Программный и математический аппарат для анализа данных

6. Основы SQL

6.1 Проектирование БД

6.2 Создание структур в базе данных и написание запросов

6.3 Импорт данных из внешних источников в БД MS SQL

6.4 Основы работы с конструктором таблиц и команд языка SQL

Государственный экзамен проводится в сочетании устной и письменной форм в формате собеседований.

Государственный экзамен проводится в два этапа:

Этап 1. Тестирования (кол-во вопросов в тесте: 30),

Этап 2. Практические задания: 2 (две) устных и 1 (одна) письменная задача.

Профессиональная область (дисциплины) для практических заданий определяется по итогам тестирования. Профильным считается область (дисциплина), по которой студент имеет наибольший процент правильных ответов по тесту. В случае равномерного распределения процента правильных ответов студент выбирает профиль на свое усмотрение.

Порядок проведения экзамена:

- выдача студенту тестового задания, фиксирования времени выдачи;
- выполнения теста в течении 30 минут, после чего ожидание вне аудитории проверки тестовой части;
- проверка теста и выдача студенту практических заданий по профессиональной области (дисциплине);
- выполнение студентом устных практических заданий (не более 30 минут на 2 устных практических заданий);
- выполнение задания студентом письменного практического задания, время выполнения – не более 1,5 часов;
- выполненное практическое задание сдается комиссии в виде:
 - скриптов (программных сценариев);
 - инструкций запуска к ним.

Скрипты должны быть подготовлены таким образом, чтобы комиссия могла запустить их выполнение без присутствия студента. Допускается выполнение всех задач в рамках одного скрипта.

- комиссией проверяются выполненные задания студентов и выставляется оценка. Время проверки всех практических заданий экзаменуемых – не более 1 часа, а студенты находятся вне аудитории до окончания проверки;

- оглашение итогового решения комиссии студентам в день проведения экзамена.

Во время экзамена разрешается пользоваться:

- подготовленным файлом requirements.txt;
- подготовленными самописными модулями;
- клонировать репозитории github;
- инструментами импорта и установки библиотек из интернета.

Во время экзамена не разрешается:

- импортировать библиотеки полностью выполняющие задания;
- использовать LLM для решения задачи.

2.2. Перечень вопросов, выносимых на экзамен.

Приведенные ниже вопросы являются примерными.

Тестовые задания

1. Машинное обучение и анализ данных

1. Вопрос. Какие задачи решают деревья решений?

- а) Только классификацию.
- б) Только регрессию.

в) Классификацию и регрессию.

г) Только анализ временных рядов.

2. Вопрос. Что такое дерево решений в контексте машинного обучения?

а) Алгоритм для классификации и регрессии, который представляет собой структуру в виде дерева, где каждый узел представляет атрибут, а листья — возможные исходы.

б) Метод анализа данных, основанный на машинном обучении, который используется для прогнозирования будущих событий.

в) Инструмент визуализации данных, позволяющий наглядно представить структуру и связи между различными элементами.

г) Статистический метод для анализа влияния различных факторов на результат.

3. Вопрос. Какой тип визуализации данных используется для демонстрации корреляции между двумя переменными?

а) Диаграмма рассеяния.

б) Гистограмма.

в) Линейный график.

г) Круговая диаграмма

4. Вопрос. Какой тип визуализации данных лучше всего подходит для отображения распределения числовых данных?

а) Столбчатая диаграмма.

б) Гистограмма.

в) Линейный график.

г) Круговая диаграмма.

5. Вопрос. Какой метод анализа данных применяется для разбиения данных на группы на основе сходства?

а) Регрессионный анализ.

б) Кластеризация.

в) Дисперсионный анализ.

г) Корреляционный анализ.

6. Вопрос. Какой метод анализа данных используется для определения взаимосвязи между двумя переменными?

а) Регрессионный анализ.

б) Кластеризация.

в) Дисперсионный анализ.

г) Корреляционный анализ.

7. Вопрос. Какое ключевое слово используется в Python для создания конструктора?

а) `__init__`.

б) `__class__`.

в) `__main__`.

г) `__name__`.

8. Вопрос. Какой метод в Python отвечает за удаление элемента списка с указанной позицией?

а) `pop()`.

б) `remove()`.

в) `delete()`.

г) `clear()`.

9. Вопрос. Какие типы JOIN существуют в SQL?

а) INNER JOIN, LEFT JOIN, RIGHT JOIN.

б) CROSS JOIN, FULL JOIN, SEMI JOIN.

в) OUTER JOIN, INNER JOIN, CROSS JOIN.

г) Все вышеперечисленные типы JOIN.

10. Вопрос. Что делает команда SELECT в SQL?

а) Создает новую таблицу.

б) Удаляет данные из таблицы.

в) Выбирает данные из таблицы.

г) Изменяет структуру таблицы.

11. Вопрос. Какой из видов машинного обучения основывается на взаимодействии обучаемой системы со средой?

а) Обучение с учителем

б) Глубинное обучение

- в) Обучение без учителя
- г) Обучение с подкреплением

12. Вопрос. Нейросети хорошо проявляют себя не только в распознавании, но и в генерации изображений. Но кое с чем у них все-таки возникают проблемы. С чем именно?

- а) Цвет
- б) Форма
- в) Глубина, количество пикселей
- г) Текстуры

13. Вопрос. У машинного обучения есть ряд задач. Как называется та, что направлена на предсказание значения той или иной непрерывной числовой величины для входных данных?

- а) Переобучение
- б) Классификация
- в) Регрессия
- г) Кластеризация

14. Вопрос. Искусственные нейронные сети (ИНС) — модели машинного обучения, использующие комбинации распределенных простых операций, зависящих от обучаемых параметров, для обработки входных данных. Какого вида ИНС не существует?

- а) Импульсные
- б) Состязательные
- в) Наивные
- г) Рекуррентные

15. Вопрос. На чем основана модель случайного леса?

- а) На ансамбле деревьев
- б) На ансамбле графов
- в) На паре деревьев
- г) На матрице смежности для ансамбля графов

16. Вопрос. Что является результатом прогноза логистической регрессии?

- а) Численное значение целевой переменной
- б) Вероятность возникновения события
- в) Погрешность возникновения некоторого события
- г) Коэффициенты линейной функции, описывающей закон возникновения события

17. Вопрос. Что из перечисленного относится к технологиям ИИ?

- а) CFD-расчёты
- б) СУБД

- в) 3D-печать
- г) Обработка естественного языка

18. Вопрос. Что из перечисленного является open-source компонентом?

- а) SAS
- б) Terradata
- в) Matplotlib
- г) Power BI

19. Вопрос. Как расшифровывается ELT?

- а) Extract-List-Transform
- б) Extract-Load-Translate
- в) Extract-Load-Transform
- г) Excel-Load-Trim

20. Вопрос. Какой из перечисленных признаков НЕ является интервальным?

- а) Рост
- б) Давление
- в) Марка автомобиля
- г) Вес

21. Вопрос. из перечисленного НЕ относится к процедурам улучшения качества данных?

- а) "Золотая запись"
- б) Идентификационный анализ
- в) Бининг
- г) Извлечение

2. Алгоритмы и структуры данных

1. Вопрос: Выберите правильное рекуррентное соотношение для создания последовательности чисел Фибоначчи:

- а) $F(0) = 0, F(1) = 1, F(n) = F(n-1) + F(n-2)$
- б) $F(0) = 1, F(1) = 1, F(n) = F(n-1) + F(n-2)$
- в) $F(0) = 1, F(1) = 2, F(n) = F(n-1) + F(n-2)$

2. Вопрос: Что такое итерация цикла в программировании?

- а) Процесс завершения выполнения программы.
- б. Процесс повторного выполнения блока кода до тех пор, пока выполняется заданное условие.
- в) Способ хранения данных в памяти.
- г) Метод оптимизации алгоритма для повышения его скорости.

3. Вопрос: Как называется структура, которая позволяет организовать данные в форме дерева, где каждый узел может иметь несколько дочерних узлов?

- а) Линейная структура
- б) Структура с иерархией
- в) Граф
- г) Динамический массив

4. Вопрос: Какое десятичное число соответствует восьмеричному числу 17?

- а) 13
- б) 14
- в) 15
- г) 16

5. Вопрос: Ниже дан фрагмент кода на языке Python. Что будет выведено на экран?

```
a = 17
b = 6
while a > 0:
    result = a // b
    remainder = a % b
    a -= b
    print(result, remainder)
```

- а) 2, 5
- б) 0, 6
- в) 0, 5
- г) 3, 4

3. Основы SQL

1. Вопрос. Что такое реляционные базы данных:

а) База данных, в которой информация хранится в виде двумерных таблиц, связанных между собой

б) База данных, в которой одна ни с чем не связанная таблица

в) Любая база данных - реляционная

г) Совокупность данных, не связанных между собой

2. Вопрос. Как выглядит запрос, для вывода ВСЕХ значений из таблицы Orders:

- а) `select ALL from Orders;`
- б) `select % from Orders;`
- в) `select * from Orders;`
- г) `select *.Orders from Orders;`

3. Вопрос. Какие данные мы получим из этого запроса?

```
select id, date, customer_name from Orders;
```

а) Неотсортированные номера и даты всех заказов с именами заказчиков

б) Никакие, запрос составлен неверно

в) Номера и даты всех заказов с именами заказчиков, отсортированные по первой колонке

г) Номера и даты всех заказов с именами заказчиков, отсортированные по всем колонкам, содержащим слово Order

4. Вопрос. Есть ли ошибка в запросе?

```
select id, date, customer_name from Orders where customer_name = Mike;
```

а) Запрос составлен правильно

б) Mike необходимо записать в кавычках 'Mike'

в) Нужно убрать лишние поля из запроса

г) Строчку с where поменять местами с from

5. Вопрос. Что покажет следующий запрос:

```
select * from Orders where date between '2017-01-01' and '2017-12-31'
```

а) Все данные по заказам, совершенным за 2017 год, за исключением 01 января 2017 года

б) Все данные по заказам, совершенным за 2017 год, за исключением 31 декабря 2017 года

в) Все данные по заказам, совершенным за 2017 год

г) Ничего, запрос составлен неверно

6. Вопрос. С помощью какого запроса можно получить пользователей, в email которых содержится более двух точек?

а) `SELECT * FROM users WHERE "." IN email MATCHES > 2;`

б) `SELECT * FROM users WHERE email LIKE "%.%.%.%";`

в) `SELECT * FROM users WHERE email COUNT(".") > 2`

7. Вопрос. Какой механизм в реляционных СУБД позволяет обеспечить согласованность данных, хранящихся в логически связанных таблицах?

а) Первичный ключ;

б) Внешний ключ;

в) Кластерный индекс;

8. Вопрос. Какой оператор позволяет объединить строки, имеющие одинаковые значения в заданных полях?

а) ORDER BY;

б) WHERE;

в) GROUP BY;

г) FROM;

9. Вопрос. Какой тип соединения двух таблиц позволит вернуть те строки, для которых в обеих таблицах выполняется условие соединения?

- а) LEFT JOIN;
- б) FULL JOIN;
- в) RIGHT JOIN;
- г) INNER JOIN;

10. Вопрос. Какой инструмент SQL позволяет получить результат обработки одной или нескольких строк, и подставить этот результат в текущую строку как новый столбец?

- а) Оператор GROUP BY;
- б) Команды соединения таблиц;
- в) Оконные функции;

11. Вопрос. Какой оператор возвращает уникальные строки, выводимые левым и правым входными запросами?

- а) UNION;
- б) EXCEPT;
- в) INTERSECT;

4. Анализ больших текстовых данных и текстовый поиск

1. Вопрос. Какой правильное определение модели Sequential:

- а) модель разбиения текста на отдельные токены/блоки
- б) полносвязная модель нейронной сети прямого распространения
- в) рекуррентная нейронная сеть

2. Вопрос. Что на вход получает encoder модели text2text?

- а) Векторное представление вопроса
- б) Векторное представление ответа
- в) Вектор контекста
- г) Векторное представление ответа и контекста

3. Вопрос. Какие функции активации обычно есть в слое LSTM?

- а) Сигмоид
- б) Tanh
- в) Relu
- г) units

4. Вопрос. Какой результатам соответствует преобразованию текста методом `texts_to_matrix`?

- а) [0, 0, 0, 1, 0, 0, 1, 0, 0, 1]
- б) [416, 10, 3]
- в) Словарь частот

5. Как задача относится к задачам ML без учителя

- а) NER

- б) Topic Model
- в) Text2Text
- г) Суммаризация
- 6. Функция активации Relu это
 - а) Линейный выпрямитель
 - б) Моделирование векторного пространства
 - в) Функция, отключающая нейроны по схеме бернулли
- 5. Основы веб-программирования
 1. Вопрос. Как расшифровывается аббревиатуру HTML?
 - а) HyperText Markup Language
 - б) HeadText Modulation Language
 - в) HeadingText Mode Language
 2. Вопрос. Где следует располагать тег title?
 - а) В теге: head
 - б) В теге: body
 - в) В теге: footer
 3. Вопрос. Существует ли у тега input закрывающий тег в синтаксисе HTML?
 - а) Да
 - б) Нет
 - в) Такого тега в стандартах HTML не существует
 4. Вопрос. Какого вида селектора не существует?
 - а) Селектор псевдокласса
 - б) Сестринский селектор
 - г) Братский селектор
 5. Вопрос. Какие ключевые слова используются в JavaScript для того, чтобы объявить переменную?
 - а) var, let, const
 - б) byte, short, int, long, float
 - в) int, short, var
 6. Вопрос. Что будет, если значению длины непустого массива присвоить ноль?
(let arr = ['test', '123']; arr.length = 0;) - если требуется пример
 - а) Ошибка
 - б) Массив станет пустым
 - в) Ничего не произойдет
 7. Вопрос. Верно ли, что на самом деле в JS всё является объектом?
 - а) Да
 - б) Нет

Практические задания

1. Алгоритмы и структура данных, python

Устные задания

1. Сколько всего куриц на планете?
2. Сколько деревьев в Москве?
3. Сколько теннисных мячиков поместится в автобус?

Письменные задания

1. Реализуйте алгоритм Дейкстры.
2. Напишите программу, которая в последовательности натуральных чисел определяет минимальное число, оканчивающееся на 3.
3. Учитывая целое число n , напишите функцию для вычисления n -го числа Фибоначчи

2. Основы SQL

Устные задания

1. Чем отличается left join от inner join
2. Чем отличаются union от union all
3. Есть две таблицы данных:

table1

Id value

1 A

2 B

2 B

3 C

table2

Id Value2

1 a

2 b

3 c

Сколько строк вернет запрос

```
Select * from table1 inner join table2 on table1.Id = table2.ID
```

4. Есть две переменные

A = null

B = null

Что вернет запрос:

IF A=B

Письменное задание

1. Есть некий скрипт sql. Что в нем происходит.
2. Вывести идентификаторы пользователей и последнюю пару вопросов-ответ по каждому пользователю, где есть текстовый ответ ассистента.

3. Напишите SQL-запрос, чтобы из таблицы изъять имена, которые начинаются на букву А.

3. Анализ больших текстовых данных и текстовый поиск

Устные задания:

1. Что такое большая лингвистическая модель?
2. Каковы компоненты архитектуры LLM?
3. Какие типы LLM вы знаете?

Письменные задания:

1. Разработать API-сервис, принимающий вопросы пользователей и выполняющий обращения к языковым моделям с применением подхода RAG. В качестве дополненной информации для RAG предполагается внутренняя Wiki база по тех. поддержке условной организации в виде XML-файлов.

Реализация сервиса:

- Векторизация данных из XML файлов происходит при запуске сервиса. Файлы берутся из папки `./source_data/`. Можно использовать любую удобную модель для функции embedding'a.

- Сохранение полученных векторов в БД. Можно использовать любую БД, подходящую для хранения векторных данных.

- При успешном запуске на стартовой странице должна быть отражена информация «Сервер запущен. Дата и Время запуска». Если во время заполнения векторной БД возникли ошибки, то информация «Сервер запущен. Локальные данные не загружены»

- Прием вопроса от пользователя происходит на странице «<http://сервер/llm>» через GET запрос:

* Параметр «query»

- Происходит поиск в векторной базе.

- Данные запроса и поиска передаются на генерацию в языковую модель.

Может быть использована любая модель как локальная так и сетевая (Пример: Gigachat API для физических лиц).

- Ответ LLM в формате { “answer”: ответ LLM } отображается на странице «<http://сервер/llm>».

Тестирование

Проект запускается из среды разработки.

Запущенный web-Сервер должен:

- Выполнить векторизацию данных в БД

- Отобразить стартовую страницу

- Принять в адресной строке вопрос пользователя (Пример: <http://сервер/llm?query=вопрос>)

- Отображать текстом ответ в формате JSON на странице <http://сервер/llm>

- Ответ должен соответствовать вопросу и содержать данные из XML файлов

Усложнение:

- Использование любых механизмов и подходов по улучшению точности выдаваемого ответа на основе прилагаемых загруженных данных (Пример: Системные промты, Перефразирование вопроса, Гибридный поиск, Переранжирование и т.д.)

Дополнительные требования:

- Доп. требования: Использование версии Python 3.9 или выше. Информация в XML файлах на русском языке. Учитывайте это при выборе моделей

- Обработка исключений: Ваш сервис должен корректно обрабатывать ошибки, например отсутствие XML файлов в папке, отсутствие доступа к LLM (если она сетевая). Ошибка должна сохраняться в текстовый файл log-дата.txt в папке `./logs/`. Один файл с логом на каждый день

- Документация: Предоставьте документацию для вашего сервиса, включая описание параметра, возвращаемое значение и любые известные ограничения, которые в него заложены на языке разметки Markdown

- Допускается использование локальных Open Source моделей и любых, общедоступных сетевых LLM с бесплатным или частично-бесплатным доступом. При использовании сетевых моделей, ключи авторизации должны быть внесены в переменные окружения.

Результат:

- Файлы с исходным кодом

- Документация по использованию функции

2. Прочитайте текстовый файл, проведите анализ тональности текста и анализ именных сущностей. Результатом выполнения программы должен быть автоматически формируемый отчет с таблицами и графическим представлением данных.

3. Прочитайте текстовый файл, напишите программу прогнозирующую количество реакций на публикацию текста в социальной сети.

4. Прочитайте текстовый файл. Проведите векторизацию текста используя различные методы, в том числе предобученные модели, word2vec, Skip-Gram

5. Прочитайте текстовый файл. Используйте его как контекст для создания диалоговой системы. Создайте диалоговые модели, используя предобученные LLM. Сформируйте отчет о результатах работы различных

моделей (сравнение ответов на одни и те же вопросы). Сделайте вывод об области применения предобученных LLM.

4. Системы программирования на Python

Устные задания:

1. В чем разница между списком и кортежем?
2. Что будет при делении числа `int` на `0`?
3. Что такое декоратор?
4. В чем разница между методами экземпляра, класса и статическими методами в Python?

Письменные задания:

1. Чтение текстового файла размером более 7 Гб и формирование единого массива данных. Время работы не должно превышать 3 минуты.

2. Напишите функцию, которая вернет пересечение двух массивов? Например, $X = [1,5,9,0]$ и $Y = [3,0,2,9]$. Функция должна учитывать, что пользователь может вводить данные любого типа, однако

3. Василий считает, что когда текст пишут в скобках (как вот тут, например), его читать не нужно. Вот и надумал он существенно укоротить время чтения, написав функцию,

которая будет удалять все, что расположено внутри скобок.

Помогите ленивому Васе разработать функцию, которая будет удалять все, что внутри скобок и сами эти скобки, возвращая очищенный текст (скобки могут быть вложенными).

4. Напишите программу для вычисления и вывода суммы квадратов двузначных чисел, которые делятся на 77 без остатка. При решении задачи используйте функции `filter()`, `map()` и `sum()`. На 77 должно делиться исходное двузначное число, а не его квадрат. Не забывайте про отрицательные двузначные числа.

```
numbers = [77, 293, 28, 242, 213, 285, 71, 286, 144, 276, 61, 298, 280, 214, 156, 227, 228, 51, -4, 202, 58, 99, 270, 219, 94, 253, 53, 235, 9, 158, 49, 183, 166, 205, 183, 266, 180, 6, 279, 200, 208, 231, 178, 201, 260, -35, 152, 115, 79, 284, 181, 92, 286, 98, 271, 259, 258, 196, -8, 43, 2, 128, 143, 43, 297, 229, 60, 254, -9, 5, 187, 220, -8, 111, 285, 5, 263, 187, 192, -9, 268, -9, 23, 71, 135, 7, -161, 65, 135, 29, 148, 242, 33, 35, 211, 5, 161, 46, 159, 23, 169, 23, 172, 184, -7, 228, 129, 274, 73, 197, 272, 54, 278, 26, 280, 13, 171, 2, 79, -2, 183, 10, 236, 276, 4, 29, -10, 41, 269, 94, 279, 129, 39, 92, -63, 263, 219, 57, 18, 236, 291, 234, 10, 250, 0, 64, 172, 216, 30, 15, 229, 205, 123, -105]
```

5. Электронные часы показывают время в формате `h:mm:ss`, то есть сначала записывается количество часов, потом обязательно двузначное количество минут, затем обязательно двузначное количество секунд.

Количество минут и секунд при необходимости дополняются до двузначного числа нулями. С начала суток прошло n секунд. Выведите, что покажут часы.

Входные данные:

Вводится целое число n .

Выходные данные:

Выведите ответ на задачу, соблюдая требуемый формат.

6. Определите класс `train` с двумя атрибутами: `type` и `speed`. Затем создайте экземпляр и верните `speed`

7. Преобразуйте следующий цикл `for` в генератор списков (`list comprehension`):

```
a = [1,2,3,4,5]
```

```
a2 = []
```

```
for i in a:
```

```
    a2.append(i + 1)
```

```
print(a2)
```

8. Напишите программу, которая проверит наличия заданной пользователем даты во фрейме данных и событие, ему соответствующее. Дата подается в формате `'2021-01-18'`. В случае если во фрейме данных даты нет, необходимо вывести оповещение и ближайшую дату.

```
data = {'Date': ['2021-01-18', '2021-01-20',
```

```
         '2021-01-23', '2021-01-25'],
```

```
        'event': ['fail', 'correct', 'fail', 'fail']}
```

9. Напишите программу, которая принимает на вход описание одного объекта в формате JSON и выводит все пары ключ-значение этого объекта.

На вход программе подается корректное описание одного объекта в формате JSON.

Программа должна вывести все пары ключ-значение введенного объекта, разделяя ключ и значение двоеточием, каждую на отдельной строке. Если значением ключа является список, то все его элементы должны быть выведены через запятую. Пары ключ-значение при выводе должны располагаться в своем исходном порядке. Для считывания произвольного числа строк используйте потоковый ввод `sys.stdin`.

5. Машинное обучение и методы обработки данных

Устные задания:

1. Какие метрики качества моделей ML вы знаете?

2. Что такое регуляризация?

3. Что такое функция потерь? Как ее можно изменить?

4. Основные этапы разработки ML-моделей

Практические задания:

1. EDA и Feature Engineering

Считайте csv-файл. Проведите разведочный анализ данных. Результаты представьте в виде:

- фрейма данных и текстового описания. Текст и фрейм должны генерироваться по итогу выполнения кода.

- Графического сопровождения, логичного для набора данных (не менее 3, исходя из данных)

2. Кластеризация

Считайте csv-файл.

Построить кластеризацию следующими методами:

- k-means (при выборе количества кластеров использовать -метод elbow)

- dbscan

Выбор переменных для кластеризации должен быть продиктован здравым смыслом и содержать не менее 5 переменных.

Построить график в параллельных координатах для обеих кластеризаций (k-means\dbscan).

Построить кластеризации kmeans\dbscan для двух и трёх переменных и показать на двумерном\трёхмерном графике результаты.

3. Anomaly Detection

Прочитайте текстовый файл. Проведите анализ аномалий.

Обоснуйте выбор инструмента выявления аномалий, проведите анализ аномалий, обоснуйте возможность использования всего объема данных для решения задач машинного обучения. В случае если это не возможно, разделите исходные данные на выборки.

4. Классификация

Прочитайте текстовый файл.

Необходимо построить базовую модель классификации, используя библиотеку для AutoML (библиотека любая - FEDOT, auto-sklearn, LightAutoML или любую другую). Выбрать (и обосновать выбор) основную метрику качества для решения задачи классификации - AUC ROC, Accuracy, F1-score и т.д. Построить для базовой модели Confusion Matrix, ROC-Curve, посчитать основную метрику качества. Построить модель классификации на базе изученных алгоритмов (деревья, лес, бустинг, логистическая регрессия и др.). Построить для вашей модели Confusion Matrix, ROC-Curve, посчитать основную метрику качества.

5. Регрессия

Прочитайте текстовый файл.

Необходимо построить базовую модель регрессии, используя библиотеку для AutoML (библиотека любая - FEDOT, auto-sklearn,

LightAutoML или любую другую). Выбрать (и обосновать выбор) основную метрику качества для решения задачи классификации - MAE, MSE, RMSE и т.д. Построить для базовой модели посчитать основную метрику качества и построить график реальных и предсказанных значений (scatter plot, heatmap). Построить модель регрессии на базе изученных алгоритмов (деревья, лес, бустинг, линейная регрессия, нейросети и др.). Построить для вашей модели график реальных и предсказанных значений, посчитать основную метрику качества.

6. Основы web-программирования

Устные задания:

1. Предназначение HTML, структура гипертекста, правила написания тегов HTML

2. Структура веб-страницы, теги и содержание мета-данных. Структура HTML тегов.

3. Принципы интеграции HTML с каскадными таблицами стилей и кодом JavaScript

4. Технология разработки веб-приложений и сайтов Модель-Представление-Контроллер (MVC), причины необходимости применения и предназначение каждого блока.

MVC (Model-View-Controller) — это архитектурный шаблон для разработки программного обеспечения, особенно веб-приложений и сайтов.

5. Структура каркаса (фреймворка) разработки веб-приложений и сайтов на языке программирования Python (Django)

6. Принципы создания динамических HTML страниц при использовании Python (Django), методы интеграции HTML с серверной частью веб-сайта.

7. Организация данных веб-сайта через модели на стороне сервера, взаимодействие веб-сайта с базами данных.

Практические задания:

1. Реализовать методы, которые в процессе выполнения строки $(2).plus(3).minus(1)$ дали бы на выходе 4.

2. Сверстать макет

В шапке сайта изображение заполняет весь блок и располагается по центру. Контент страницы может быть от 1 до 1000 строк. Количество пунктов и подпунктов правого меню может варьироваться в диапазоне от 1 до 10. Эффект для разворачивания подменю выбирается самостоятельно. Боковое меню является фиксированным при прокрутке до тех пор, пока блок соцсетей не оказывается на расстоянии 50 пикселей от футера. При этом следует учитывать различные размеры меню, контента и экрана.

3. Создать компонент для отображения таблицы с данными (Таблица с фейковыми данными из JSON'а, данные любого вида)

Получить данные с backend'а и вывести полученные данные в табличку, описание запросов было предоставлено. Кнопка добавления новой записи. Запись посылаем на бэкэнд. Кнопка удаления записи напротив каждой строки таблицы, по клику запись удаляется запросом на бэкэнде, по успеху — на фронтэнде удаляется

Кнопка редактирования напротив каждой строки таблицы. По клику на нее, в таблице все инпуты становятся редактируемыми и на месте кнопки редактировать, появляется кнопка сохранить. По нажатию на сохранить, отправляется запрос на бэкэнд, по успеху — инпуты опять становятся ридонли и кнопка меняется на «редактировать».

2.3. Рекомендации обучающимся по подготовке к экзамену.

Подготовка к экзамену способствует закреплению, углублению и обобщению знаний, получаемых, в процессе обучения, а также применению их к решению практических задач. Готовясь к государственному экзамену, студент ликвидирует имеющиеся пробелы в знаниях, углубляет, систематизирует и упорядочивает свои знания. На государственном экзамене студент демонстрирует то, что он приобрел в процессе обучения. В период подготовки к государственному экзамену студенты вновь обращаются к учебно-методическому материалу и закрепляют знания. Подготовка студента к государственному экзамену включает в себя три этапа: самостоятельная работа в течение всего периода обучения; непосредственная подготовка в дни, предшествующие государственному экзамену по темам разделам и темам учебных дисциплин, выносимых на государственную аттестацию. При подготовке к государственному экзамену студентам целесообразно использовать материалы лекций, методические материалы образовательной программы, рекомендованные правовые акты, основную и дополнительную литературу.

В процессе экзаменационного ответа преподавателем оценивается не только знание того или иного вопроса, но и ряд других, не лежащих на поверхности факторов к числу которых, в первую очередь, относится культура гуманитарного знания, профессиональное оперирование терминологией, культура речи студента. Во время ответа на поставленные вопросы надо быть готовым к дополнительным или уточняющим вопросам. Дополнительные вопросы задаются членами государственной комиссии в рамках билета и связаны, как правило, с неполным ответом. Уточняющие вопросы задаются,

чтобы либо конкретизировать мысли студента, либо чтобы студент подкрепил те или иные теоретические положения практикой, либо привлек знания смежных учебных дисциплин. Полный ответ на уточняющие вопросы лишь усиливает эффект общего ответа студента. Итоговая оценка знаний предполагает дифференцированный подход к студенту, учет его индивидуальных способностей, степень усвоения и систематизации основных теоретических положений, понятий и категорий. Оценивается так же культура речи, грамотное комментирование, приведение примеров, умение связывать теорию с практикой, творчески применять знания к неординарным ситуациям, излагать материал доказательно, подкреплять теоретические положения знанием нормативных актов, полемизировать там, где это необходимо.

2.4. Перечень рекомендуемой литературы

№ п/п	Библиографическое описание	Место доступа
1	<p>Рашка, С. Python и машинное обучение: крайне необходимо пособие по новейшей предсказательной аналитике, обязательно для более глубокого понимания методологии машинного обучения : руководство / С. Рашка ; перевод с английского А. В. Логунова. — Москва : ДМК Пресс, 2017. — 418</p>	<p>https://e.lanbook.com/book/100905?category_pk=1556&publisher__fk=1028&ysclid=1w60n980di451663507</p>

	с. — ISBN 978-5-97060- 409-0	
2	<p>Йылдырым, С. Осваиваем архитектуру Transformer / С. Йылдырым, М. Асгари- Ченаглу ; перевод с английского В. С. Яценкова. — Москва : ДМК Пресс, 2022. — 320 с. — ISBN 978-5-93700- 106-1</p>	https://e.lanbook.com/book/241148
3	<p>Алетдинова, А. А. Интеллектуа льный анализ больших данных : учебное пособие / А. А. Алетдинова, М. Ш. Муртазина. — Новосибирс к : НГТУ, 2023. — 66 с. — ISBN 978-5-7782- 4899-1</p>	https://e.lanbook.com/book/404567
4	<p>Павлов, Л. А. Структуры и</p>	<a href="https://e.lanbook.com/book/156929?category=1540&ysclid=lw4yv0ho
hi198485137">https://e.lanbook.com/book/156929?category=1540&ysclid=lw4yv0ho hi198485137

	<p>алгоритмы обработки данных : учебник для вузов / Л. А. Павлов, Н. В. Первова. — 3-е изд., стер. — Санкт-Петербург : Лань, 2021. — 256 с. — ISBN 978-5-8114-7259-8</p>	
5	<p>Пантелеев, Е. Р. Алгоритмы и структуры данных : учебное пособие / Е. Р. Пантелеев, А. Л. Алыкова. — Иваново : ИГЭУ, 2018. — 142 с.</p>	<p>https://e.lanbook.com/book/154576</p>
6	<p>Таблицы интегралов и другие математические формулы Г.Б. Двайт; Ред. К.А. Семендяев; Пер. с англ. Н.В. Леви Однотомное издание "Лань" , 2005</p>	<p>НТБ (уч.6); НТБ (фб.); НТБ (чз.2)</p>

7	<p>Ростовцев, В. С. Искусственные нейронные сети / В. С. Ростовцев. — 4-е изд., стер. — Санкт-Петербург : Лань, 2024. — 216 с. — ISBN 978-5-507-47362-5</p>	<p>https://e.lanbook.com/book/364517?ysclid=m1rz4vi0fi416423386</p>
8	<p>Соробин, А. Б. Сверточные нейронные сети: примеры реализаций : учебно-методическое пособие / А. Б. Соробин. — Москва : РТУ МИРЭА, 2020. — 159 с.</p>	<p>https://e.lanbook.com/book/163853?category=1547&ysclid=m1rz5u3q5x476258334</p>
9	<p>Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль ; перевод с английского А. А. Слинкина. — 2-е изд. — Москва :</p>	<p>https://e.lanbook.com/book/107901</p>

	<p>ДМК Пресс, 2018. — 652 с. — ISBN 978-5-97060- 618-6</p>	
1 0	<p>Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах. — Москва : ДМК Пресс, 2015. — 400 с. — ISBN 978-5-97060- 273-7</p>	<p>https://e.lanbook.com/book/69955</p>
1 1	<p>Голицына, О. Л. Базы данных : учебное пособие / О.Л. Голицына, Н.В. Максимов, И.И. Попов. — 4-е изд., перераб. и доп. — Москва : ФОРУМ : ИНФРА-М, 2020. — 400 с. — (Среднее профессиона льное образование</p>	<p>https://znanium.ru/catalog/document?id=362825&ysclid=lweo1o32q4992160167#bib</p>

	<p>). - ISBN 978-5-00091- 601-8</p>	
1 2	<p>Карпова, Т. С. Базы данных: модели, разработка, реализация : учебное пособие / Т. С. Карпова. — 2-е изд. — Москва : ИНТУИТ, 2016. — 403 с.</p>	<p>https://e.lanbook.com/book/100575?category_pk=1556&ysclid=lweo4bpupg644993495</p>
1 3	<p>Габдуллин, Н. М. Развитие человеческо го капитала и цифровой экономики в регионах России: факторный и кластерный анализ : монография / Н. М. Габдуллин. — Казань : КФУ, 2019. — 268 с. — ISBN 978-5- 00130-291-9</p>	<p>https://e.lanbook.com/book/173018</p>
1 4	<p>Гласснер, Э. Глубокое обучение без математики. Том 2. Практика : руководство / Э.</p>	<p>https://e.lanbook.com/book/131710</p>

	<p>Гласснер ; перевод с английского В. А. Яроцкого. — Москва : ДМК Пресс, 2020. — 610 с. — ISBN 978-5-97060- 767-1</p>	
1 5	<p>Гультяева, Т. А. Методы статистичес кого обучения в задачах регрессии и классификац ии : монография / Т. А. Гультяева, А. А. Попов, А. С. Саутин. — Новосибирс к : НГТУ, 2016. — 323 с. — ISBN 978-5-7782- 2817-7</p>	<p>https://e.lanbook.com/book/118291</p>
1 6	<p>Кук, Д. Машинное обучение с использован ием библиотеки H2O / Д. Кук ; перевод с английского А. Б. Огурцова. — Москва :</p>	<p>https://e.lanbook.com/book/97353</p>

	<p>ДМК Пресс, 2018. — 250 с. — ISBN 978-5-97060- 508-0</p>	
1 7	<p>Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах. — Москва : ДМК Пресс, 2015. — 400 с. — ISBN 978-5-97060- 273-7</p>	<p>https://e.lanbook.com/book/69955</p>
1 8	<p>Шалев- Шварц, Ш. Идеи машинного обучения : учебное пособие / Ш. Шалев- Шварц, Бен- Давид Ш. ; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2019. — 436 с. — ISBN 978-5-97060- 673-5</p>	<p>https://e.lanbook.com/book/131686</p>
1 9	<p>Гудфеллоу, Я. Глубокое</p>	<p>https://e.lanbook.com/book/107901</p>

	<p>обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль ; перевод с английского А. А. Слинкина. — 2-е изд. — Москва : ДМК Пресс, 2018. — 652 с. — ISBN 978-5-97060- 618-6</p>	
2 0	<p>Болотова, Ю. А. Методы и алгоритмы интеллекту альной обработки цифровых изображени й : учебное пособие / Ю. А. Болотова, А. А. Друки, В. Г. Спицын. — Томск : ТПУ, 2016. — 208 с. — ISBN 978-5- 4387-0710-3</p>	<p>https://e.lanbook.com/book/107751</p>
2 1	<p>Шапиро, Л. Компьютерн ое зрение : учебное пособие / Л. Шапиро, Д. Стокман ; под редакцией</p>	<p>https://e.lanbook.com/book/135496</p>

	<p>С. М. Соколова ; перевод с английского А. А. Богуславско го. — 4-е изд. — Москва : Лаборатория знаний, 2020. — 763 с. — ISBN 978-5-00101- 696-0</p>	
2 2	<p>Нуньес- Иглесиас, Х. Элегантный SciPy / Х. Нуньес- Иглесиас, в. д. Уолт, Х. Дэшноу. — Москва : ДМК Пресс, 2018. — 266 с. — ISBN 978-5-97060- 600-1</p>	<p>https://e.lanbook.com/book/116124</p>
2 3	<p>Волосова, А. В. Технологии искусственн ого интеллекта в ULS- системах / А. В. Волосова. — 2-е изд., испр. — Санкт- Петербург : Лань, 2024. — 308 с. —</p>	<p>https://e.lanbook.com/book/370217</p>

	ISBN 978-5-507-45885-1	
2 4	Ненашев, В. А. Компьютерное зрение. Анализ, обработка и моделирование : учебное пособие / В. А. Ненашев. — Санкт-Петербург : ГУАП, 2022. — 78 с. — ISBN 978-5-8088-1806-4	https://e.lanbook.com/book/341057

3. Перечень компетенций, которые должны быть сформированы у обучающихся в результате освоения образовательной программы.

ОПК-2 - Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности;

ОПК-3 - Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности;

ОПК-5 - Способен устанавливать программное и аппаратное обеспечение для информационных и автоматизированных систем;

ОПК-8 - Способен разрабатывать алгоритмы и программы, пригодные для практического применения;

ПК-1 - Способен анализировать большие данные с использованием существующей в организации методологической и технологической инфраструктуры;

ПК-4 - Способен выполнять работы и управлять работами по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы.

4. Критерии оценки результатов итоговой (государственной итоговой) аттестации.

4.1. Критерии оценки результатов сдачи экзамена.

Шкала оценивания	Критерии
Оценивание результатов 1	Правильно выполнено 90% и более заданий (отсутствие ответа приравнивается к неверно выполненному заданию). На очной части экзамена: - Задание выполнено полностью, - ответ обоснованно, - выводы и предложения аргументированы и оформлены должным образом Выпускник демонстрирует навыки публичной дискуссии, защиты собственных научных и/или практических идей, предложений и рекомендаций, культура общения с аудиторией на высоком уровне.
Оценивание результатов 1	Приведено свыше 80-89% правильных ответов из тестовых заданий (отсутствие ответа приравнивается к неверно выполненному заданию). На очной части экзамена: - задание выполнено полностью, но допущены незначительные неточности в расчетах или оформлении; - или при условии надлежащего оформления, задания, выполнено не меньше, как на 80%
Оценивание результатов 1	Приведено свыше 70-89% правильных ответов из тестовых заданий (отсутствие ответа приравнивается к неверно выполненному заданию). На очной части экзамена: - задание выполнено не менее как на 70% при условии надлежащего оформления; или не менее как на 80% при условии незначительных ошибок в расчетах
Оценивание результатов 1	Приведено менее 55% правильных ответов из тестовых заданий (отсутствие ответа приравнивается к неверно выполненному заданию). На очной части экзамена: - задание выполнено менее как на 50%, допущены принципиальные ошибки в расчетах

Авторы:

О.Б. Проневич

Согласовано:

Директор

Б.В. Игольников

Председатель учебно-методической
комиссии

Д.В. Паринов