

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЛУГАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ ВЛАДИМИРА ДАЛЯ»

Краснодонский факультет инженерии и менеджмента (филиал)
Кафедра информационных технологий и транспорта



УТВЕРЖДАЮ:
Директор
Панайотов К.К.

«14» марта 2025 года

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по учебной дисциплине

Анализ данных средствами языка программирования Python

(наименование учебной дисциплины, практики)

38.04.05 Бизнес-информатика

(код и наименование направления подготовки (специальности))

«Бизнес-аналитика»

наименование профиля подготовки (специальности, магистерской программы); при отсутствии ставится прочерк)

Разработчик(разработчики):

доцент

Бихдрикер А.С.

(подпись)

ФОС рассмотрен и одобрен на заседании кафедры информационных технологий и транспорта от «26» февраля 2025 г., протокол № 7

Заведующий кафедрой
информационных
технологий и транспорта

Верительник Е.А

(подпись)

Краснодон 2025

**Комплект оценочных материалов по дисциплине
«Анализ данных средствами языка программирования Python»**

Задания закрытого типа

Задания закрытого типа на выбор правильного ответа

Выберите один правильный ответ.

1. Какая библиотека Python наиболее часто используется для работы с табличными данными, такими как CSV-файлы или данные из баз данных?

- A) matplotlib.
- Б) numpy.
- В) pandas.
- Г) scikit-learn.

Правильный ответ: В

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

2. *Выберите один правильный ответ.*

Какая функция из библиотеки numpy используется для создания многомерного массива?

- A) array().
- Б) matrix().
- В) arange().
- Г) linspace().

Правильный ответ: А

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

3. Какая библиотека Python чаще всего используется для визуализации данных в виде графиков и диаграмм?

- A) seaborn.
- Б) numpy.
- В) pandas.
- Г) scikit-learn.

Правильный ответ: А

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

4. Какая библиотека Python используется для построения моделей машинного обучения?

- A) matplotlib.
- Б) numpy.
- В) pandas.
- Г) scikit-learn.

Правильный ответ: Г

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

Задания закрытого типа на установление соответствие

Установите правильное соответствие.

Каждому элементу левого столбца соответствует только один элемент правого столбца.

1. Установите соответствие между библиотекой Python и её назначением.

Библиотека Python	Назначение
1) pandas	A) Визуализация данных, создание графиков и диаграмм.
2) numpy	Б) Работа с многомерными массивами и выполнение математических операций.
3) matplotlib	В) Работа с табличными данными, анализ и манипуляция данными.
4) seaborn	Г). Визуализация данных на основе matplotlib, но с более высокой абстракцией и красивыми стилями.

Правильный ответ: 1В, 2Б, 3А, 4Г

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

2. Сопоставьте функции pandas и их действия.

Функция	Действие
1) read_csv()	А) Удаление строк или столбцов с пропущенными значениями.
2) groupby()	Б) Заполнение пропущенных значений в DataFrame.
3) fillna()	В) Чтение данных из CSV-файла в DataFrame.
4) dropna()	Г) Группировка данных в DataFrame по одному или нескольким столбцам.

Правильный ответ: 1В, 2Г, 3Б, 4А

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

3. Сопоставьте типы графиков и их предназначение.

Тип графика	Предназначение
1) Гистограмма	А) Отображение взаимосвязи между двумя переменными.
2) Диаграмма рассеяния (Scatter plot)	Б) Отображение частотного распределения данных.
3) Столбчатая диаграмма (Bar chart)	В) Сравнение значений категориальных данных.

Тип графика	Предназначение
4) Линейный график (Line plot)	Г) Отображение изменения данных во времени или по другой непрерывной переменной.

Правильный ответ: 1Б, 2А, 3Г, 4Д

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

4. Сопоставьте термины машинного обучения и их значения.

Термин	Значение
1) Переобучение (Overfitting)	А) Состояние, когда модель слишком сложна и хорошо работает на тренировочных данных, но плохо на новых данных.
2) Недообучение (Underfitting)	Б) Состояние, когда модель слишком проста и плохо работает как на тренировочных, так и на новых данных.
3) Конструирование признаков (Feature Engineering)	В) Процесс выбора, преобразования и комбинирования признаков для улучшения производительности модели.
4) Оценка модели (Model Evaluation)	Г) Процесс оценки качества обученной модели на тестовых данных.

Правильный ответ: 1А, 2Б, 3В, 4Г

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

Задания закрытого типа на установление правильной последовательности

Установите правильную последовательность.

Запишите правильную последовательность букв слева направо.

1. Процесс создания визуализации данных с использованием Matplotlib:

- А) Добавление заголовка к графику.
- Б) Настройка осей (например, добавление подписей).
- В) Выбор типа графика (например, столбчатая диаграмма, линейный график).
- Г) Отображение графика.
- Д) Подготовка данных для графика.

Правильный ответ: Д, В, Б, А, Г

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

2. Процесс создания и оценки модели машинного обучения с использованием Scikit-learn.

- А) Оценка производительности модели на тестовом наборе данных.
- Б) Разделение данных на обучающий и тестовый наборы.

В) Выбор модели машинного обучения (например, линейная регрессия, дерево решений).

Г) Обучение модели на обучающем наборе данных.

Д) Предварительная обработка данных (например, масштабирование признаков).

Правильный ответ: Д, Б, В, Г, А

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

3. Процесс анализа данных с использованием pandas и matplotlib.

А) Визуализация данных с помощью matplotlib.

Б) Импорт библиотеки pandas.

В) Чтение данных из файла (например, CSV) с помощью pandas.

Г) Предварительная обработка данных (очистка, преобразование).

Д) Анализ данных с использованием pandas (например, группировка, агрегация).

Правильный ответ: Б, В, Г, Д, А

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

4. Обработка пропущенных значений в pandas DataFrame.

А) Проверка наличия пропущенных значений в DataFrame.

Б) Выбор стратегии обработки пропущенных значений (например, удаление или заполнение).

В) Импорт библиотеки pandas.

Г) Чтение данных в pandas DataFrame.

Д) Применение выбранной стратегии для обработки пропущенных значений.

Правильный ответ: В, Г, А, Б, Д

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

Задания открытого типа

Задания открытого типа на дополнение

Напишите пропущенное слово (словосочетание).

1. Метод _____ в pandas позволяет объединить два DataFrame по общему столбцу или индексу, что полезно для интеграции данных из разных источников.

Правильный ответ: merge() / merge.

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

2. Для расчета корреляции между столбцами в pandas DataFrame используется метод _____, который позволяет оценить степень линейной зависимости между переменными.

Правильный ответ: corr() / corr.

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

3. Функция _____ из библиотеки `scipy.stats` позволяет провести t-тест для проверки гипотезы о равенстве средних значений двух выборок. Результаты теста, как правило, включают t-статистику и p-значение.

Правильный ответ: `ttest_ind / ttest_ind()`.

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

4. Чтобы оценить, насколько хорошо модель машинного обучения (например, линейная регрессия) объясняет дисперсию зависимой переменной, используется метрика под названием _____, принимающая значения от 0 до 1.

Правильный ответ: R-квадрат / R^2 / Коэффициент детерминации.

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

Задания открытого типа с кратким свободным ответом

Дайте ответ на вопрос.

1. Какой метод в `pandas` позволяет получить основные статистические характеристики `DataFrame`, такие как среднее значение, стандартное отклонение и медиану?

Правильный ответ: `describe() / Описание / Статистика / describe.`

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

2. Как называется метод в `pandas`, позволяющий отфильтровать строки `DataFrame` по определенному условию?

Правильный ответ: `query() / Запрос / Выборка / Фильтрация / query.`

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

3. Какое понятие в машинном обучении описывает ситуацию, когда модель слишком хорошо адаптируется к обучающим данным и плохо обобщает на новые данные?

Правильный ответ: Переобучение / Overfitting / Переучивание.

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

4. Как называется процесс в анализе данных, при котором данные преобразуются в формат, более подходящий для моделирования?

Правильный ответ: Feature Engineering / Конструирование признаков / Инженерия признаков / Создание новых признаков.

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

Задания открытого типа с развернутым ответом

Дайте развернутый ответ на вопрос.

1. Опишите, как бы вы подошли к задаче выбора наиболее подходящих признаков для модели машинного обучения в Python. Какие методы вы бы использовали и почему? Объясните разницу между фильтрацией, оберткой и встроенными методами выбора признаков

Время выполнения - 20 мин.

Ожидаемый результат:

Выбор признаков – это критический шаг для повышения производительности и интерпретируемости модели. Начать необходимо с понимания задачи и данных. Затем применить методы:

1. Фильтрация. Основана на статистических показателях (например, корреляции с целевой переменной). Быстрая, но не учитывает взаимодействие признаков. Примеры: выбор признаков на основе коэффициента корреляции Пирсона.

2. Обертка. Оценивает подмножества признаков, обучая модель на каждом подмножестве. Требует больших вычислительных затрат. Примеры: рекурсивное исключение признаков (RFE).

3. Встроенные. Используют алгоритмы, которые автоматически выполняют выбор признаков в процессе обучения. Примеры: L1-регуляризация в линейных моделях.

Разница: фильтрация – независима от модели, обертка – использует модель для оценки подмножеств, встроенные – являются частью алгоритма. Выбор зависит от размера данных и вычислительных ресурсов. Часто используют комбинацию методов.

Критерии оценивания: Наличие в ответе двух методов и объяснение разницы между ними.

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

2. Объясните концепцию регуляризации в моделях машинного обучения. Какие типы регуляризации существуют и как они помогают бороться с переобучением?

Время выполнения - 15 мин.

Ожидаемый результат:

Регуляризация – это метод, используемый для предотвращения переобучения моделей машинного обучения путем добавления штрафа к сложным моделям.

Типы регуляризации.

1. L1 (Lasso) – добавляет штраф, пропорциональный абсолютной величине коэффициентов. Может привести к обнулению некоторых коэффициентов, что влияет на выбор признаков.

2. L2 (Ridge) – добавляет штраф, пропорциональный квадрату величины коэффициентов. Уменьшает величину коэффициентов, но не обнуляет их.

3. Elastic Net – Комбинация L1 и L2 регуляризации.

Регуляризация помогает бороться с переобучением, упрощая модель и уменьшая ее зависимость от конкретных обучающих данных.

Критерии оценивания: Каждый тип регуляризации должен быть описан минимум одним предложением.

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

3. Опишите, как вы бы использовали Python и библиотеки pandas и scikit-learn для построения модели логистической регрессии для задачи классификации. Какие метрики можно использовать для оценки качества модели?

Время выполнения - 20 мин.

Ожидаемый результат:

Логистическая регрессия используется для задач классификации (прогнозирование принадлежности к одному из классов).

Этапы построения модели логистической регрессии для задачи классификации.

1. Подготовка данных. Загрузка данных, обработка пропущенных значений, кодирование категориальных признаков.
2. Разделение данных. Разделение на обучающий и тестовый наборы.
3. Обучение модели. Создание объекта LogisticRegression и обучение его на обучающем наборе.
4. Прогнозирование. Прогнозирование классов на тестовом наборе.

Метрики оценки.

1. Accuracy – доля правильно классифицированных объектов.
2. Precision – доля верно определенных объектов положительного класса среди всех объектов, определенных как положительные.
3. Recall – доля верно определенных объектов положительного класса среди всех объектов положительного класса.
4. F1-score – гармоническое среднее между precision и recall.
5. AUC-ROC – площадь под кривой ROC, характеризующая способность модели различать классы.

Критерии оценивания: Каждый этап должен быть описан минимум одним предложением и наличие в ответе трёх метрик оценки.

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

4. Объясните разницу между «обучающим» (training), «проверочным» (validation) и «тестовым» (test) наборами данных в машинном обучении. Зачем нужны эти наборы и как их правильно применять?

Время выполнения - 15 мин.

Ожидаемый результат:

Для построения надежной модели машинного обучения, важно разделить доступные данные на три части.

Обучающий набор – используется для обучения модели, то есть для настройки её параметров.

Проверочный набор – используется для настройки гиперпараметров модели и для выбора наилучшей модели из нескольких обученных вариантов. Позволяет избежать переобучения.

Тестовый набор – используется для окончательной оценки производительности обученной модели. Модель оценивается на тестовом наборе только один раз, после того как выбраны лучшие гиперпараметры и лучшая модель.

Правильное применение. Сначала модель обучается на обучающем наборе, затем гиперпараметры оптимизируются на проверочном наборе. Наконец, финальная модель оценивается на тестовом наборе, чтобы получить объективную оценку её обобщающей способности. Перемешивание данных перед разделением важно.

Критерии оценивания: Описание всех наборов данных и принцип их применения.

Компетенции (индикаторы): ОПК-5 (ОПК-5.2), ПК-2 (ПК-2.2)

Экспертное заключение

Представленный фонд оценочных средств (далее – ФОС) по дисциплине «Численные методы в экономических расчетах» соответствует требованиям ФГОС ВО.

Предлагаемые формы и средства текущего и промежуточного контроля адекватны целям и задачам реализации основной профессиональной образовательной программы по направлению подготовки 38.03.05 Бизнес-информатика.

Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины представлены в полном объеме.

Виды оценочных средств, включенные в представленный фонд, отвечают основным принципам формирования ФОС.

Разработанный и представленный для экспертизы фонд оценочных средств рекомендуется к использованию в процессе подготовки обучающихся по указанному направлению 38.03.05 Бизнес-информатика.

Председатель учебно-методической комиссии Краснодонского факультета инженерии и менеджмента (филиала)

Ольга Родионова О.Ю.

Лист изменений и дополнений

№ п/п	Виды дополнений и изменений	Дата и номер протокола заседания кафедры (кафедр), на котором были рассмотрены и одобрены изменения и дополнения	Подпись (с расшифровкой) заведующего кафедрой (заведующих кафедрами)