

**«Финансовый университет при Правительстве Российской Федерации»
(Финуниверситет)**

Калужский филиал Финуниверситета

Кафедра «Бизнес-информатика и высшая математика»

«УТВЕРЖДАЮ»

Директор Калужского филиала

Финуниверситета



В.А. Матчинов

27 июня 2024 г.

Костенко А.В.

**ПРОГРАММИРОВАНИЕ И АНАЛИЗ ДАННЫХ
С ПОМОЩЬЮ PYTHON**

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

для студентов, обучающихся по направлению подготовки

38.03.02 «Менеджмент»

Образовательная программа бакалавриата: «Управление бизнесом»

профиль «Менеджмент и управление бизнесом»

очная форма обучения

*Рекомендовано Ученым советом Калужского филиала Финуниверситета
(протокол № 16 от 27.06. 2024 г.)*

Одобрено кафедрой «Бизнес – информатика и высшая математика»
Калужского филиала Финуниверситета
(протокол № 12 от 27 июня 2024 г.)

КАЛУГА 2024


Рабочая программа предназначена для преподавания дисциплины «Программирование и анализ данных с помощью Python» студентам, обучающимся по направлению подготовки 38.03.02 «Менеджмент», образовательная программа бакалавриата: «Управление бизнесом» профиль «Менеджмент и управление бизнесом» по очной форме обучения.

В рабочей программе излагаются планируемые результаты освоения дисциплины, содержание дисциплины, тематика и содержание семинаров и практических занятий, технологии их проведения. В рабочей программе дисциплины приводится перечень учебно-методического обеспечения для самостоятельной работы обучающихся, фонд оценочных средств для проведения промежуточной аттестации обучающихся, перечень основной и дополнительной литературы, а также ресурсов информационно-телекоммуникационной сети Интернет.

За основу взята программа М. В. Смирнов Программирование и анализ данных с помощью Python Рабочая программа дисциплины. Финуниверситет. Москва 2023.

СОГЛАСОВАНО:


Заместитель директора
по учебно-методической работе
«27» июня 2024 г.

 /Орловцева О.М./

Начальник учебно-методического отдела
«27» июня 2024 г.

 /Толстикова В.С./

Заведующий кафедрой
«Бизнес-информатика и высшая математика»
«27» июня 2024 г.

 /Дробышева И.В./

Содержание

| | | |
|------|---|----|
| 1. | Наименование дисциплины | 4 |
| 2. | Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине | 4 |
| 3. | Место дисциплины в структуре образовательной программы | 5 |
| 4. | Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся | 5 |
| 5. | Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий | 5 |
| 5.1 | Содержание дисциплины | 5 |
| 5.2 | Учебно-тематический план | 8 |
| 5.3 | Содержание семинаров, практических занятий | 9 |
| 6. | Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине | 11 |
| 6.1 | Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы | 11 |
| 6.2 | Перечень вопросов, заданий, тем для подготовки к текущему контролю | 12 |
| 7. | Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине | 15 |
| 8. | Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины | 20 |
| 9 | Перечень ресурсов информационно – телекоммуникационной сети «Интернет», необходимых для освоения дисциплины | 20 |
| 10. | Методические указания для обучающихся по освоению дисциплины | 21 |
| 11. | Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем | |
| 11.1 | Комплект лицензионного программного обеспечения | 23 |
| 11.2 | Современные профессиональные базы данных и информационные справочные системы | 23 |
| 11.3 | Сертифицированные программные и аппаратные средства защиты информации | 23 |
| 12. | Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине | 23 |

1. Наименование дисциплины

«Программирование и анализ данных с помощью Python».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Таблица 1

| Код компетенции | Наименование компетенции | Индикаторы достижения компетенции | Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции |
|-----------------|---|---|---|
| ПКН-2 | Способность применять математические методы для решения стандартных профессиональных задач, интерпретировать полученные математические результаты. | 1. Демонстрирует знания математических методов, применяемых в менеджменте. | Знать: возможности информационных технологий для выбора приоритетов в процессе планирования работ, связанных с обработкой и анализом данных. Уметь: выбирать методы и способы их реализации на языке программирования Пайтон для принятия решений и совершенствования профессиональной деятельности. |
| | | 2. Применяет математические методы и модели для обоснования принятия управленческих решений. | Знать: библиотеки и модули языка программирования Пайтон для проведения научных исследований, обработки и анализа данных. Уметь: использовать язык программирования Пайтон для применения методов обработки и анализа данных в рамках решения практических и научных задач. |
| | | 3. Содержательно интерпретирует результаты, полученные при использовании математических моделей. | Знать: методы и модели анализа данных, применяемые в профессиональной деятельности. Уметь: проводить анализ и вырабатывать собственное профессиональное мнение на основе имеющихся данных. |
| ПКП-4 | Способность участвовать в разработке программ развития компании, разработке обоснований проектов и управленческих решений, связанных с развитием бизнеса. | 1. Разрабатывает концепцию проекта, иерархическую структуру работ, календарно-ресурсный план и контроль за ходом программ развития организации. | Знать: состав и структуру данных для хранения и обработки на языке программирования Пайтон. Уметь: применять язык программирования Пайтон для реализации методов обработки и анализа данных различного состава и структуры. |
| | | 2. Применяет современные модели развития и управления организацией. | Знать: возможности языка программирования Пайтон в выявлении и моделировании причинно-следственных связей признаков наборов данных. Уметь: создавать программы на языке Пайтон для выявления |

| | | | |
|--|--|--|--|
| | | | закономерностей в анализируемых данных и оценки степени их изменчивости. |
|--|--|--|--|

3. Место дисциплины в структуре образовательной программы

Дисциплина «Программирование и анализ данных с помощью Python» имеет порядковый номер Б.1.1.3.2. в структуре основной образовательной программы по направлению подготовки 38.03.02 – Менеджмент, ОП «Управление бизнесом» профиль «Менеджмент и управление бизнесом» и относится к циклу профиля (элективный).

4. Объем дисциплины(модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

Таблица 2

| Вид учебной работы по дисциплине | Часы: | |
|---|------------------|------------|
| | Всего | Сем. 3 |
| Общая трудоёмкость дисциплины | 108 | 108 |
| Контактная работа - Аудиторные занятия | 36 | 36 |
| Лекции | 2 | 2 |
| Семинары, практические занятия | 34 | 34 |
| Самостоятельная работа | 72 | 72 |
| Вид текущего контроля | проектная работа | |
| Вид промежуточной аттестации | зачет | |

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Введение в программирование на языке Python

История создания, области применения, отличительные признаки языка программирования *Python*. Версии языка программирования *Python*.

Знакомство с программным обеспечением *Jupyter Notebook*. Создание нового документа *Jupyter Notebook*, типы ячеек, особенности написания кода в ячейках. Язык разметки *Markdown*, написание формул и приёмы оформления работ с помощью *Markdown*.

Облачные сервисы, реализующие интерактивную вычислительную среду *IPython Notebook*.

Задание и использование переменных, арифметические операторы. Библиотеки языка Python, их версии и обновление. Библиотека *math*.

Тема 2. Типы данных и управляющие конструкции

Числовые типы данных языка *Python*.

Функции и методы работы со строками. Итерация символов строки, срезы. Поиск подстроки по шаблону.

Операторы сравнения. Логические переменные и операторы. Условные операторы. Циклы *while* и *for*.

Тема 3. Коллекции

Работа со списками в Python. Создание списка. Операции над списками. Перебор элементов списка. Методы списков.

Работа со словарями в Python. Создание словаря. Операции над словарями. Перебор элементов словаря. Методы работы со словарями.

Работа с кортежами в Python. Операции и методы обработки данных в кортежах.

Работа с множествами в Python. Создание множества, методы работы с множествами.

Тема 4. Функции

Определение функции. Аргументы функции, переменное число аргументов, аргументы по умолчанию, последовательность аргументов.

Локальные и глобальные переменные. Функция как объект. Анонимная функция *lambda*. Функции *map()* и *filter()*. Рекурсивные функции.

Организация хранения и использования функций в модулях.

Тема 5. Обработка и анализ данных в *NumPy*-массивах

Библиотека *NumPy*: установка, назначение, использование. *NumPy*-массивы:

создание, размерность, оси, индексация, срезы. Типы данных элементов массива. Копия и представление *NumPy*-массива. Объединение, укладка, разделение массивов. Поиск в массиве. Сортировка массива. Фильтрация массива. Случайные числа в *NumPy*-массиве. Функции библиотеки *NumPy* для обработки данных в *NumPy*-массивах. Распространение, маскирование и прихотливое индексирование.

Тема 6. Обработка и анализ табличных данных с помощью *Pandas*

Библиотека *Pandas*: установка, назначение, особенности использования. Типы данных *DataFrame* и *Series*. Индексация табличных данных, метки строк и столбцов. Обработка данных вдоль соответствующей оси с помощью функции *apply()*, описательная статистика данных в строках и столбцах.

Типы табличных данных, числовые и категориальные признаки. Форматирование даты и преобразование типов.

Извлечение данных по условию. Группировка данных. Сводные таблицы. Объединение таблиц.

Тема 7. Методы визуализации данных

Библиотеки *Matplotlib* и *Seaborn*. Цветовые схемы, маркеры. Отображение на графике нескольких показателей. Различные виды графиков, используемых для визуализации результатов анализа данных: столбчатая диаграмма, гистограмма распределения признаков, линейный график, точечная диаграмма, ящик с усами.

Объектно-ориентированный подход к созданию графиков. Область построения графика и система координат как объекты. Основной и вспомогательный графики в одной системе координат, несколько графиков в одной области построения.

Тема 8. Обработка и анализ данных в файлах

Структурированные и неструктурированные данные. Работа с файлами, чтение и запись данных в файлах, бинарное и текстовое представление данных, юникод.

Обработка и анализ документов или сообщений в форматах CSV, JSON. XML, XLSX, HTML. Библиотеки для обработки структурированных данных: *csv*, *json*, *ElementTree*, *xlwings*.

5.2 Учебно-тематический план

Таблица 3

| п/п | Наименование тем (разделов) дисциплины | Трудоемкость в часах | | | | Формы текущего контроля успеваемости | |
|-----|--|----------------------|--|--------|---------------------------------------|---|---------------------------|
| | | Всего | Контактная работа – Аудиторная работа | | | | Самостоятельная работа |
| | | | Общая, в т. ч.: | Лекции | Семинары, практическ ие занятия | | |

| | | | | | | | |
|---|---|-----|----|---|----|----|---|
| 1 | Введение в программирование на языке Python | 12 | 4 | 2 | 2 | 8 | Активная работа и решение задач на семинарах, выполнение домашних заданий |
| 2 | Типы данных и управляющие конструкции | 12 | 4 | - | 4 | 8 | |
| 3 | Коллекции | 14 | 4 | - | 4 | 10 | |
| 4 | Функции | 14 | 6 | - | 6 | 8 | Активная работа и решение задач на семинарах, выполнение домашних заданий |
| 5 | Обработка и анализ данных в <i>NumPy</i> массивах | 14 | 4 | - | 4 | 10 | |
| 6 | Обработка и анализ табличных данных с помощью <i>Pandas</i> | 14 | 4 | - | 4 | 10 | |
| 7 | Методы визуализации данных | 14 | 4 | - | 4 | 10 | Активная работа и решение задач на семинарах, выполнение домашних заданий |
| 8 | Обработка и анализ данных в файлах | 14 | 6 | - | 6 | 8 | |
| | В целом по дисциплине | 108 | 36 | 2 | 34 | 72 | Согласно учебному плану: проектная работа |
| | Итого в % | | 33 | 6 | 94 | 67 | |

5.3 Содержание семинаров, практических занятий

Таблица 4

| Наименование тем (разделов) дисциплины | Перечень вопросов для обсуждения на семинарах, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника) | Формы проведения занятий |
|--|---|--------------------------|
|--|---|--------------------------|

| | | |
|---|---|--|
| Тема 1. Введение в программирование на языке Python | Интерфейс <i>Jupyter Notebook</i> , установка, обновление, особенности программирования. Переменные, арифметические операции. <i>Источники: 8.1, 8.2, 8.3, 9.1, 9.3.</i> | Индивидуальное выполнение заданий, групповой разбор результатов (30% времени на интерактивные технологии) |
| Тема 2. Типы данных и управляющие конструкции | Изучение базовых конструкций языка программирования Python. Строки, методы работы со строками. Поиск в строке. Циклы. <i>Источники: 8.2, 8.4, 9.2, 9.3.</i> | Индивидуальное выполнение заданий, групповой разбор результатов (30% времени на интерактивные технологии) |
| Тема 3. Коллекции | Списки, словари, кортежи, множества. Определение, создание, наполнение, итерация, срезы. Выражения-генераторы. <i>Источники: 8.1, 8.2, 9.2, 9.3.</i> | Индивидуальное выполнение заданий, групповой разбор результатов (30% времени на интерактивные технологии) |
| Тема 4. Функции | Определение функции. Аргументы, локальные и глобальные переменные. Функция как объект. Анонимная функция <i>lambda</i> . Функции <i>map()</i> и <i>filter()</i> . Рекурсивные функции. <i>Источники: 8.2, 9.2, 9.3.</i> | Индивидуальное выполнение заданий, групповой разбор результатов (30% времени на интерактивные технологии) |
| Тема 5. Обработка и анализ данных в <i>NumPy</i> массивах | Библиотека <i>NumPy</i> , <i>NumPy</i> -массивы: размерность, индексация, срезы. Форма массива. Объединение, укладка, разделение массивов. Поиск в массиве. Сортировка массива. Фильтрация массива. Случайный массив, статистические функции. Распространение. <i>Источники: 8.1, 8.2, 8.3, 9.2, 9.3, 9.13.</i> | Решение задач в интерактивной форме, дискуссия по результатам самостоятельной работы (30% времени на интерактивные технологии) |
| Тема 6. Обработка и анализ табличных данных с помощью <i>Pandas</i> | Библиотека <i>Pandas</i> : методы <i>describe()</i> и <i>info()</i> . Тип данных «серия». Создание таблиц <i>pandas</i> из коллекций. Типы данных в столбцах. Сводные таблицы, статистические функции. Применение функций к столбцам. <i>Источники: 8.1, 8.3, 9.2, 9.3, 9.14.</i> | Решение задач в интерактивной форме, дискуссия по результатам самостоятельной работы (30% времени на интерактивные технологии) |
| Тема 7. Методы визуализации данных | Графический функционал библиотеки <i>Pandas</i> . Объектно-ориентированный подход к созданию графиков. Виды графиков. Библиотека <i>Matplotlib</i> . <i>Источники: 8.1, 9.2, 9.3, 9.15.</i> | Решение задач в интерактивной форме, дискуссия по результатам самостоятельной работы (30% времени на интерактивные технологии) |
| Тема 8. Обработка и анализа данных в файлах | Библиотеки Python для обработки структурированных данных: <i>csv</i> , <i>json</i> , <i>ElementTree</i> , <i>xlwings</i> . Обработка и анализ документов или сообщений в форматах <i>CSV</i> , <i>JSON</i> , <i>XML</i> , <i>XLSX</i> . <i>Источники: 9.2, 9.3, 9.16, 9.18, 9.19.</i> | Решение задач в интерактивной форме, дискуссия по результатам самостоятельной работы (30% времени на интерактивные технологии) |

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

При изучении дисциплины используются следующие виды самостоятельной работы:

- Выполнение внеаудиторной проектной работы.
- Подготовка к зачету.
- Работа с литературой: проработка учебного материала (по конспектам лекций, учебной и научной литературе).
- Самостоятельное изучение тем теоретического курса, не вошедших в лекционный материал.
- Работа с тестами и вопросами для самопроверки: выполнение тренировочных тестов.
- Подготовка к интерактивному занятию.
- Подготовка к тестированию.

Таблица 5

| Наименование тем (разделов) дисциплины | Перечень вопросов, отводимых на самостоятельное освоение | Формы внеаудиторной самостоятельной работы |
|---|---|---|
| Тема 1. Введение в программирование на языке Python | Использование языка разметки <i>Маркдаун</i> для оформления научных работ. Горячие клавиши Юпитер Ноутбук. Альтернативные интерактивные оболочки. | Изучение интерфейса и возможностей Юпитер ноутбук. Написание математических формул на языке <i>Маркдаун</i> . |
| Тема 2. Типы данных и управляющие конструкции | Десятичное, двоичное и шестнадцатеричное представление чисел в Python. Побитовые операторы. | Чтение учебных изданий и материалов в Интернете, выполнение упражнений в <i>Jupyter Notebook</i> . |
| Тема 3. Коллекции | Вложенные коллекции: списки, словари, кортежи, множества. Создание коллекций с помощью выражений-генераторов. | Чтение учебных изданий и материалов в Интернете, выполнение упражнений в <i>Jupyter Notebook</i> . |
| Тема 4. Функции | Организация хранения и использования функций в модулях. Функция <i>reduce()</i> . | Чтение учебных изданий и материалов в Интернете, выполнение упражнений в <i>Jupyter Notebook</i> . |

| | | |
|---|--|--|
| Тема 5. Обработка и анализ данных в <i>NumPy</i> массивах | Создание <i>NumPy</i> -массивов на основе данных, хранящихся в файлах. | Чтение учебных изданий и материалов в Интернете, выполнение упражнений в <i>Jupyter Notebook</i> . |
| Тема 6. Обработка и анализ табличных данных с помощью <i>Pandas</i> | Представление данных экономической тематики из открытых источников в табличном виде. Расчёт описательной статистики. | Изучение открытых источников данных, разработка программ в <i>Jupyter Notebook</i> . |
| Тема 7. Методы визуализации данных | Трехмерные графики. Маркеры, цветовые схемы. | Работа с <i>Matplotlib</i> , графическое представление информации. |
| Тема 8. Обработка и анализа данных в файлах | Обработка открытых данных в форматах XML, CSV, JSON. | Практика с использованием финансово-экономических данных в файлах |

6.2 Перечень вопросов, заданий, тем для подготовки к текущему контролю (согласно таблице 2)

Текущий контроль осуществляется в ходе учебного процесса и контроля самостоятельной работы студентов, а также по результатам выполнения аудиторных самостоятельных работ, контрольной работы. Основными формами текущего контроля знаний являются:

- обсуждение вопросов по теме практических занятий, в том числе выступления у доски;
- выполнение домашних и аудиторных практических работ и обсуждение результатов;
- выполнение проектной работы.

Промежуточная аттестация проводится в форме зачета.

Примерные задачи семинарских занятий

1. Имеются два списка целых чисел, возможно разной длины, упорядоченные по возрастанию. Получите новый список, содержащий все элементы исходных списков, в котором элементы также упорядочены в порядке возрастания, не используя встроенные методы сортировки.

2. Какую функцию надо применить в сочетании с функцией *map()*, чтобы список слов ["Я", "студент", "Финансового", "университета"] преобразовать в список чисел [1, 7, 11, 12]? Напишите программу, реализующую эту задачу.

3. С помощью функций *filter* и *lambda* в списке [4, 8, 15, 16, 23, 42] оставьте только элементы со значениями выше среднего.

4. Найдите все числа ряда Фибоначчи меньше заданного натурального числа *n*, используя рекурсивную функцию.

5. Имеется список городов вида [["город", "год основания", "население"], ["Гомель", 1142, 504], ["Бобруйск", 1387, 210], ["Лида", 1323, 103]]. Выполните задания: а) создайте новый список, в котором первый вложенный список остаётся на своём месте, а оставшиеся вложенные списки отсортированы по населению; б) создайте словарь, в котором ключами являются элементы первого вложенного списка, а значениями элементы остальных вложенных списков; в) поместите полученный в результате выполнения пункта «б)» словарь в таблицу *pandas*, в которой названиями столбцов являются ключи, а значениями — значения словаря.

6. Участники олимпиады решали 3 задачи. Известны фамилии тех, кто решил первую, вторую и третью задачи (для каждой задачи отдельный список). Найдите и распечатайте фамилии тех, кто а) решил хотя бы одну задачу (любую); б) решил все задачи; в) решил ровно 1 задачу (любую); г) решил ровно 2 задачи (любые); д) решил не больше 2 задач (любых).

7. Имеются данные в формате JSON о множестве рецептов блюд национальной кухни. Каждый рецепт содержит множество ингредиентов. Найдите ингредиенты, входящие в состав а) одного блюда; б) одной кухни; в) максимального числа блюд.

8. Задать два двумерных массива *ar1* и *ar2* размерности (4, 7), состоящих из случайных целых чисел в пределах от 0 до 10. Построить двумерный массив размерности (4, 7), каждый элемент которого представляет собой максимум (минимум, среднее) из двух значений, находящихся на аналогичной позиции в массивах *ar1*, *ar2*.

9. Сгенерировать двумерный массив *arr* размерности (4, 7), состоящий из случайных действительных чисел, равномерно распределенных в диапазоне от 0 до 20. Нормализовать значения массива так, что после нормализации максимальный элемент массива будет равен 1.0, минимальный 0.0. Значения округлить до четырех десятичных знаков.

10. Имеются данные в формате CSV о финансово-экономической деятельности группы предприятий и содержащие сведения о типе выпускаемой продукции, наименовании продукции, цене, себестоимости, сегменте рынка, прибыли, дате поставки, размере скидки, стране. Найдите все сделки определенного производителя, поставленные в определенную страну при условии, что размер скидки ниже средней в соответствующем сегменте рынка.

11. Имеются данные в формате JSON о финансово-экономической деятельности группы предприятий: тип выпускаемой продукции, наименование продукции, цена, себестоимость, сегмент рынка, прибыль, дата поставки, размер скидки, страна. Напишите программу, результатом которой является сводная таблица, в которой расположите наименования производителей по строкам, а сегменты рынка по столбцам, поместите в область значений среднюю прибыль каждого производителя в каждом сегменте. Рассчитайте общее среднее значение прибыли по каждому производителю и по каждому сегменту.

12. Имеется реляционная база данных студентов, проживающих в разных городах и обучающихся на разных курсах в университетах, расположенных в разных городах. Выполните SQL-запрос, результатом которого станет список студентов, обучающихся в своем городе и получающих стипендию выше средней стипендии на своём курсе.

13. Имеется реляционная база данных студентов, проживающих в разных городах и обучающихся на разных курсах в университетах, расположенных в разных городах. Напишите запрос, выполняющий вывод количества экзаменов, сдававшихся (вне зависимости от оценки) студентами с идентификаторами 12, 32, 55. Результат должен содержать идентификатор студента и количество сданных им экзаменов.

Результат упорядочить по убыванию количества экзаменов.

14. Имеется файл финансово-экономических данных из открытого источника (*finam.ru*) о цене акции или какого-либо инвестиционного инструмента за период времени. Постройте с помощью языка Python временную модель данных. Отобразите данные и модель данных на графике. Оцените качество модели с помощью коэффициента детерминации. Рассчитайте математическое ожидание и стандартное отклонение. Постройте график функции плотности вероятности.

15. Имеются данные из открытого источника о ценах двух акций или какихлибо продуктов за период времени в одном из структурированных форматов: CSV, JSON, XSLX, XML. Постройте с помощью языка Python регрессионную модель данных. Постройте график модели. Оцените качество модели с помощью коэффициента детерминации.

Примерная тематика проектной работы

1. Анализ открытых данных городского сервиса проката велосипедов с целью построения распределения времени поездки и расчёта описательной статистики.

2. Декомпозиция открытых данных в текстовом виде сервиса сдачи в аренду жилых помещений, создание реляционной базы данных и ее наполнение из указанного источника. Выполнение SQL-запросов к созданной базе данных с целью поиска конкретной информации о сдаче жилья в аренду.

3. Регрессионный анализ финансово-экономических данных о стоимости акций на Московской товарно-сырьевой бирже.

4. Анализ открытых социально-экономических данных на сайте Правительства Москвы. Библиотеки, школы, парки Москвы: анализ режима работы, наличие специализированного оборудования, виды растений в парках и скверах.

5. Используя открытые данные сайта Правительства Москвы, проанализируйте данные численности населения, статистику коммунальных платежей и другие социально-экономические данные о жителях различных районов Москвы.

6. Анализ данных о пользователях и группах социальной сети ВКонтакте: ранжировать пользователей сообщества по значимости на основании числа связей с другими пользователями.

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях кафедры.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень компетенций, формируемых в процессе освоения дисциплины, содержится в разделе 2. Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения, соотнесенных с планируемыми результатами обучения по дисциплине.

Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, знаний и умений

Таблица 7

| Наименование компетенции | Индикаторы достижения компетенции | Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции | Типовые контрольные задания |
|--|--|---|---|
| ПКН-2 Способность применять математические методы для решения стандартных профессиональных задач, интерпретировать полученные математические результаты. | 1. Демонстрирует знания математических методов, применяемых в менеджменте. | Знать: возможности информационных технологий для выбора приоритетов в процессе планирования работ, связанных с обработкой и анализом данных. Уметь: выбирать методы и способы их реализации на языке программирования Пайтон для принятия решений и совершенствования профессиональной деятельности. | <i>Данные из открытых источников</i> представьте в табличном виде и рассчитайте ковариацию признаков. <i>На основе имеющихся данных</i> постройте точечные диаграммы для каждой пары признаков. Между какими признаками можно предположить наличие зависимости? |
| | 2. Применяет математические методы и модели для обоснования принятия управленческих решений. | Знать: библиотеки и модули языка программирования Пайтон для проведения научных исследований, обработки и анализа данных. Уметь: использовать язык программирования Пайтон для применения методов обработки и анализа данных в рамках решения | <i>Постройте гистограммы</i> всех показателей и опишите их: являются ли они симметричными или скошенными, похожи ли они на «колокол» нормального распределения? <i>Выполните</i> регрессионный анализ показателей в представленном источнике данных и оцените качество |

| | | | |
|--|---|--|---|
| | | практических и научных задач. | регрессионной модели. Постройте точечную диаграмму |
| | 3. Содержательно интерпретирует результаты, полученные при использовании математических моделей. | Знать: методы и модели анализа данных, применяемые в профессиональной деятельности. Уметь: проводить анализ и вырабатывать собственное профессиональное мнение на основе имеющихся данных. | <i>Определите признаки</i> , значимые для классификации анализируемого процесса или явления, оцените полноту результатов классификации. <i>Проанализируйте</i> табличные данные и выделите группы данных, значимые для проведения исследования. |
| ПКП-4 Способность участвовать в разработке программ развития компании, разработке обоснований проектов и управленческих решений, связанных с развитием бизнеса. | 1. Разрабатывает концепцию проекта, иерархическую структуру работ, календарно-ресурсный план и контроль за ходом программ развития организации. | Знать: состав и структуру данных для хранения и обработки на языке программирования Пайтон. Уметь: применять язык программирования Пайтон для реализации методов обработки и анализа данных различного состава и структуры. | <i>Выполните запрос</i> к информационной системе социальной сети и получите данные о пользователях и группах. Ранжируйте членов сообщества по значимости. |
| | 2. Применяет современные модели развития и управления организацией. | Знать: возможности языка программирования Пайтон в выявлении и моделировании причинно-следственных связей признаков наборов данных. Уметь: создавать программы на языке Пайтон для выявления закономерностей в анализируемых данных и оценки степени их изменчивости. | Проведите десериализацию открытого набора данных «Что приготовить» («What's cooking», https://www.kaggle.com/c/whatscooking/data). Для каждого ингредиента найдите число блюд, в которые он входит. |

Примерные вопросы для подготовки к зачету

1. Встроенные числовые типы языка Python.
2. Списки. Создание, основные операции.
3. Основные методы списка.
4. Кортежи. Создание, основные методы и операции.
5. Словари. Создание, основные операции. Методы для работы со словарями.
6. Множества. Создание, основные методы и операции.
7. Переменные. Правила именования переменных.
8. Динамическая типизация.
9. Операторы сравнения и логические операторы.
10. Инструкция if...else.
11. Инструкция цикла while.
12. Инструкция цикла for.
13. Создание и вызов функции.
14. Передача аргументов функцию.

15. Функции-генераторы.
16. Лямбда-функции.
17. Модули. Инструкции import и from.
18. Модуль datetime. Операции с датой.
19. Работа с датой и временем с помощью strptime() и strftime().
20. Методы работы со строками. Срезы строк.
21. Методы работы со списками. Генератор списков.
22. Методы работы со словарями. Итерация по элементам словаря.
23. Методы работы с множествами. Итерация по элементам множества.
24. Методы работы с кортежами.
25. Логические операторы OR, AND. Условные выражения If ... Elif ... Else.
26. Цикл for, метод enumerate() и функция range().
27. Цикл while. Выражения break, continue, else.
28. Различные алгоритмы сортировки: их достоинства и недостатки на различных наборах данных.
29. Numpy-массивы: определение, типы данных, форма, размерность.
30. Numpy-массивы: распространение.
- 31.

Примерные задания для подготовки к зачету

1. Используя генератор списков (и не используя код вне него) преобразуйте строку по следующей логике: для каждого символа исходной строки создайте в итоговом списке строку, содержащую копии символа в количестве, равном номеру символа, рассчитанному с конца исходной строки. Пример: 'abcd' -> ['aaaa', 'bbb', 'cc', 'd']
2. Поменяйте местами столбцы с индексами 1 и 2 в массиве np.arange(9).reshape(3,3).
3. Дан массив: c = np.array([[1,2], [2,3], [4,5]]), Чему равен срез c[:,1]?
4. Пользователь вводит три числа через пробел. Присвойте их объектам a, b, n. Посчитайте и выведите на экран сумму всех чисел, кратных n, расположенных в интервале от a - 1 до b + 1 включительно.
5. Напишите функцию для вычисления факториала: а) в цикле; б) с помощью рекурсивной функции.
6. Рекурсивно реализовать функцию fib(n) вычисляющую значение n-го числа Фибоначчи. Числа Фибоначчи — элементы числовой последовательности, в которой каждое последующее число равно сумме двух предыдущих чисел (Числа Фибоначчи: 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, ...).
7. Инвертировать последовательность слов, разделенных запятыми. Пример:

строка «один, два, три, четыре, пять» будет преобразована в: «пять, четыре, три, два, один».

8. На основе строки, представляющей из себя предложение, построить вложенный список, содержащий символы всех слов в предложении. Пример: строка

```
[['В', 'о', 'л', 'ш', 'е', 'б', 'н', 'о', 'й'],  
 ['к', 'р', 'а', 'с', 'о', 'т', 'ы'],  
 ['з', 'а', 'к', 'а', 'т'],  
 ['о', 'з', 'а', 'р', 'и', 'л'],  
 ['К', 'р', 'а', 'с', 'н', 'о', 'я', 'р', 'с', 'к']]
```

«Волшебной красоты закат озарил Красноярск» должна быть преобразована в:

Рис. 1. Результат работы программы

9. В строке "Финансовый университет" удалите пробелы, и создайте словарь, в котором ключами являются чётные номера символов получившейся после удаления пробелов строки. Считайте, что первый символ строки имеет номер 0, второй символ номер 1 и так далее. Распечатайте чётные номера и соответствующие им символы.

Ожидается результат {0: "Ф", 2: "н", 4: "н", ..., 20: "т"}

10. Напишите функцию, возвращающую наибольшую степень натурального числа k , не превышающую данное натуральное число n .

11. Напишите функцию, которая находит максимум функции $f(x)$ в точках отрезка $[a, b]$ с постоянным шагом h . Параметрами функции являются f , a , b , h . Параметры a , b , h – необязательные, по умолчанию $a=0$, $b=1$, $h=0.1$. Используя эту функцию, найдите максимум функции $(2-x)\sin(x/2)$ на отрезке $[0, 4]$.

12. Напишите функцию, которая находит наибольший общий делитель (НОД) двух чисел, используя модифицированный алгоритм Евклида: нужно заменять большее число на остаток от деления большего на меньшее до тех пор, пока этот остаток не станет равен нулю; тогда меньшее число и есть НОД. Функция должна возвращать найденное значение.

13. Создайте функцию с параметрами h и p , которая рисует равнобедренный треугольник высоты h символами p .

14. Напишите функцию, преобразующую прописные буквы в строчные, удаляющую знаки препинания и возвращающую строку в обратном порядке с первой прописной буквой. Пример: Строка "Дорога, идущая в гору" в результате работы функции преобразуется в строку "Урог в ящуди агород".

15. Напишите функцию, принимающую один аргумент – год, и возвращающую True, если год високосный, иначе False.

16. Напишите функцию подсчета суммы главной диагонали матрицы. Если матрица не является квадратной, происходит выброс исключения.

17. Дан список S состоящий из N различных элементов. Вывести индексы четных элементов списка. Использовать функции высшего порядка.

18. Найдите точку $M = (x, y)$ максимума функции $f(x) = x \cdot \sin(x - 1)$ в интервале $[-\pi, \pi]$ с шагом $s = 0,1$.

19. Даны целые положительные числа m и n . Сформировать целочисленный список списков размера $m \times n$, у которого все элементы j -го столбца имеют значение $3 \cdot j$.

20. Создайте матрицу из 20 случайных целых чисел от 0 до 100. Получите второе сверху значение в матрице. Определите индекс этого значения.

21. Создайте матрицу 8 на 10 (используя модуль `numpy.random`) из случайных целых чисел в интервале от 0 до 10 и найдите в ней строку (ее индекс и вывести саму строку), в которой сумма значений минимальна.

22. Создайте матрицу размера $n \times m$ элементами которой являются дни недели, случайно выбранные из списка. Полученную матрицу сохраните в файл `matrix.csv`.

23. Используя генератор списков преобразовать список, содержащий положительные целые числа в список, элементами которого являются списки с длиной равной соответствующему числу в первом списке. Содержимым вложенных списков являются последовательно идущие целые числа начиная с 1. Пример: [3, 1, 4] -> [[1, 2, 3], [1], [1, 2, 3, 4]].

24. Распечатайте текущую дату в следующем порядке «Сегодня: День = ..., Месяц = ..., Год =».

25. Создайте генератор списков, возвращающий три первых символа каждого члена итерируемого объекта. Пример: имеется кортеж `products=("хурма", "халва", "мандарин")`, генератор списков возвращает список ["хур", "хал", "ман"].

26. Имеется список названий месяцев: ['января', 'февраля', 'марта', 'апреля', 'мая', 'июня', 'июля', 'августа', 'сентября', 'октября', 'ноября', 'декабря']. Создайте по этому списку словарь, в котором название месяца будет ключом, а номер месяца (от 1 до 12) – значением. Используя полученный словарь преобразуйте строку с датой вида «1 января 2016» в строку «1.01.2016».

27. Используя генератор словарей (и не используя код вне) инвертировать словарь, т. е. сделать ключи словаря, его значениями и наоборот. Значения, которые в исходном словаре повторяются не добавлять в итоговый словарь. Пример: {'a':1, 'b':3, 'c':4, 'd':3} -> {1:'a', 4:'c'}

28. Даны целые числа $N > 2$, a и b . Сформировать и вывести список размера N , первый элемент которого равен a , второй равен b , а каждый последующий элемент равен сумме всех предыдущих.

29. Имеются два массива $A = \text{np.array}([1,2,3,2,3,4,3,4,5,6])$ и $B = \text{np.array}([7,2,10,2,7,4,9,4,9,8])$. Получите индексы, для которых элементы массивов A и B совпадают.

30. Задайте два двумерных массива $ar1$ и $ar2$ размерности (4, 7), состоящих из случайных целых чисел в пределах от 0 до 10. Постройте двумерный массив размерности (4, 7), каждый элемент которого представляет собой максимум (минимум, среднее) из двух значений, находящихся на аналогичной позиции в массивах $ar1$, $ar2$.

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература:

1. Криволапов, С. Я. Математика на Python : учебник / С. Я. Криволапов, М. Б. Хрипунова. — Москва : КноРус, 2022. — 455. — ЭБС BOOK.ru. - URL: <https://book.ru/book/943665> (дата обращения: 15.06.2023). — Текст : электронный.

Дополнительная литература:

2. Криволапов, С. Я. Статистические вычисления на платформе Jupyter Notebook с использованием Python: учебник / С. Я. Криволапов. — Москва : КноРус, 2022. — 431 с. — ЭБС BOOK.ru. - URL: <https://book.ru/book/943660> (дата обращения: 15.06.2023). — Текст : электронный.
3. Мясникова, Н. А. Алгоритмы и структуры данных: учебное пособие / Н. А. Мясникова. — Москва : КноРус, 2023. — 185 с. — ЭБС BOOK.ru. - URL: <https://book.ru/book/946265> (дата обращения: 15.06.2023). — Текст : электронный.
4. Дадян, Э. Г. Данные: хранение и обработка: учебник по направлениям подготовки бакалавриата / Э. Г. Дадян; Финуниверситет. — Москва : Инфра-М, 2021. - 205 с. - Текст : непосредственный. - То же. - ЭБС ZNANIUM.com. - URL: <https://znanium.com/catalog/product/1149101> (дата обращения: 15.06.2023). - Текст : электронный.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Документация по проекту Юпитер (Jupyter Project Documentation): <https://docs.jupyter.org/en/latest/>
2. Официальный сайт языка программирования Пайтон <https://python.org>
3. Электронная библиотека Финансового университета: <http://library.fa.ru/>
4. Научная электронная библиотека eLibrary: <http://elibrary.ru>
5. Информационная система СПАРК: <https://spark-interfax.ru/>
6. Портал открытых данных Правительства Москвы: <https://data.mos.ru/>
7. Портал открытых данных Российской Федерации: <https://data.gov.ru/>
8. Официальный сайт SQLite. <https://sqlite.org/>
9. Сайт с учебными материалами по SQL: <https://www.sqltutorial.org/>
10. Учебный ресурс SQLAlchemy Tutorial. <https://www.tutorialspoint.com/sqlalchemy/index.htm>
11. Официальный сайт SQLAlchemy: <https://www.sqlalchemy.org/>
12. Сайт компании Финам <https://www.finam.ru/>
13. Документация библиотеки NumPy <https://numpy.org/>
14. Документация <https://pandas.pydata.org/pandas-docs/stable/index.html>
15. Документация библиотеки Matplotlib <https://matplotlib.org/>
16. Python для Excel и Гугл таблиц: <https://www.xlwings.org/>
17. Проект beautifulsoup4 <https://pypi.org/project/beautifulsoup4/>
18. CSV-модуль языка Python <https://docs.python.org/3/library/csv.html>

19. Модули для обработки XML – XML Processing Modules, <https://docs.python.org/3/library/xml.html>
20. Модуль для тестирования программ *pytest*, <https://docs.pytest.org/en/7.2.x/>
21. Объектно-ориентированное программирование на Python, <https://realpython.com/python3-object-oriented-programming/>
22. Электронно-библиотечная система Znanium <http://www.znaniy.com>
23. Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>
24. Электронно-библиотечная система издательства «Лань» <https://e.lanbook.com/>
25. Электронный учебный курс Поколение Python: курс для продвинутых. <https://stepik.org/course/68343/promo>
26. Базовый электронный учебный курс Алгоритмы и структуры данных в языке Python». <https://campus.fa.ru/course/view.php?id=20758>

10. Методические указания для обучающихся по освоению дисциплины

Обучающимся в рамках самостоятельной работы следует использовать Методические рекомендации по планированию и организации внеаудиторной самостоятельной работы студентов по образовательным программам бакалавриата и магистратуры в Финансовом университете, утвержденные Приказом ректора №1040/о от 11.05.2021 г.

Проведение семинарских и практических занятий осуществляется в компьютерных классах и включает в себя работу с различными программными продуктами и интернет-сервисами.

При изучении теоретического материала необходимо опираться на рабочую программу дисциплины, материалы лекций и литературу из основного и дополнительного списков.

Так как в данной рабочей программе предусмотрено время на самостоятельное изучение дисциплины, студентам предлагаются задания для самостоятельной работы. Методика проведения практических занятий заключается в совместном решении студентами под руководством преподавателя типовых задач и кейсов по изучаемым темам дисциплины.

Результаты выполнения заданий студенты хранят в личной сетевой папке в компьютерной сети вуза и, по требованию преподавателя, направляют преподавателю на проверку, используя корпоративные адреса электронной почты.

Проектная работа выполняется на семинарском занятии в группах (командах) по 6-8 студентов. Каждая команда выполняет аналогичную задачу на своём наборе данных, каждая команда может выбирать свои инструменты,

методы и подходы. Преподаватель определяет цель, задачи и критерии оценки результатов проектной работы.

Учитывая пройденные темы дисциплины, стандартными задачами проектной работы могут выступать задачи машинного обучения, финансово-экономическая задача по обработке рядов данных цены акций или задача проектирования и наполнения реляционной базы данных на основе данных из открытых источников.

В ходе проектной работы студенты загружают открытые данные из Интернета, строят графики и диаграммы, моделируют имеющиеся данные с помощью методов корреляционно-регрессионного анализа, при этом используют полученные знания о работе с коллекциями и библиотеками Matplotlib, NumPy и Pandas. Команды планируют и выполняют проектную работу за отведенное время. По итогам выполнения работы студенты выступают с кратким (3-5 минут) докладом о достигнутых результатах.

Методические рекомендации по обучению лиц с ограниченными возможностями здоровья

Профессорско-преподавательский состав знакомится с психологофизиологическими особенностями обучающихся инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ). При необходимости осуществляется дополнительная поддержка преподавания тьюторами, психологами, социальными работниками, прошедшими подготовку ассистентами.

В соответствии с методическими рекомендациями Минобрнауки РФ (утв. 8 апреля 2014 г. № АК-44/05вн) в курсе предполагается использовать социальноактивные и рефлексивные методы обучения, технологии социокультурной реабилитации с целью оказания помощи в установлении полноценных межличностных отношений с другими студентами, создании комфортного психологического климата в студенческой группе. Подбор и разработка учебных материалов производятся с учётом предоставления материала в различных формах: аудиальной, визуальной, с использованием специальных технических средств и информационных систем. Освоение дисциплины лицами с ОВЗ осуществляется с использованием средств обучения общего и специального назначения (персонального и коллективного использования). Материально-техническое обеспечение предусматривает приспособление аудиторий к нуждам лиц с ОВЗ. Форма проведения аттестации для студентов-инвалидов устанавливается с учётом индивидуальных психофизических особенностей.

Для студентов с ОВЗ предусматривается доступная форма предоставления заданий оценочных средств, а именно:

- в печатной или электронной форме (для лиц с нарушениями опорнодвигательного аппарата);
- в печатной форме или электронной форме с увеличенным шрифтом и контрастностью (для лиц с нарушениями слуха, речи, зрения);
- методом чтения ассистентом задания вслух (для лиц с нарушениями зрения). Студентам с инвалидностью увеличивается время на подготовку ответов

на контрольные вопросы.

Для таких студентов предусматривается доступная форма предоставления ответов на задания, а именно:

- письменно на бумаге или набором ответов на компьютере (для лиц с нарушениями слуха, речи);
- выбором ответа из возможных вариантов с использованием услуг ассистента (для лиц с нарушениями опорно-двигательного аппарата);
- устно (для лиц с нарушениями зрения, опорно-двигательного аппарата).

При необходимости для обучающихся с инвалидностью процедура оценивания результатов обучения может проводиться в несколько этапов.

Методические рекомендации по выполнению проектной работы

Методические рекомендации по выполнению проектной работы предусмотрены в «Методических рекомендациях по подготовке написанию и оформлению проектной работы», разрабатываемой преподавателем кафедры на учебный год, в котором реализуется учебная дисциплины

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

11.1 Комплект лицензионного программного обеспечения

1. Антивирусная защита ESET NOD32
2. Операционная система Astra Linux
3. Программное обеспечение Jupyter Notebook
4. Офисные программы Либре Офис

11.2 Современные профессиональные базы данных и информационные справочные системы

1. Информационно-правовая система «Консультант Плюс»;
2. Информационно-правовая система «Гарант»;

11.3 Сертифицированные программные и аппаратные средства защиты информации

Сертифицированные программные и аппаратные средства защиты информации не предусмотрены

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

- аудитории для проведения лекционных и семинарских занятий, оборудованные видеопроекционным оборудованием для презентаций, средствами звуковоспроизведения, экраном, доской меловой/интерактивной;
- библиотеку, имеющую рабочие места для студентов, оснащенные компьютерами с доступом к базам данных и сети Интернет
- компьютерные классы с набором лицензионного базового программного обеспечения для проведения практических занятий и выходом в глобальную сеть Internet;

Образовательные технологии

В процессе изучения дисциплины «Программирование и анализ данных с помощью Python» предполагается:

- сопровождение курса лекций наглядной презентацией, включающей практические примеры, схемы, графики, табличный материал;
- рассмотрение на семинарских занятиях интерактивных ситуационных задач по проблематике дисциплины;
 - деловые игры;
 - разбор конкретных ситуаций, коллективное обсуждение проблем российской и зарубежной практики по изучаемым темам;
- виртуальное общение в течение срока изучения курса в целях обеспечения лекций и практических занятий необходимым материалом и также контроля самостоятельной работы студентов.