

Федеральное государственное образовательное
бюджетное учреждение высшего образования
**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**
(Финансовый университет)

Краснодарский филиал Финуниверситета

Кафедра «Математика и информатика»

СОГЛАСОВАНО

ООО «Портал-Юг»
Генеральный директор



Е.В. Мостовой

«21» февраля 2024 г.

УТВЕРЖДАЮ

Краснодарский филиал
Финуниверситета

Директор



Э.В.Соболев

«21» февраля 2024 г.

Дюдин М.С.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
ТЕХНОЛОГИИ ОБРАБОТКИ ДАННЫХ**

студентов, обучающихся по направлению подготовки

01.03.02 Прикладная математика и информатика

в соответствии с образовательными стандартами Финансового университета
(программа подготовки бакалавров)

*Рекомендовано Ученым советом Краснодарского филиала Финуниверситета
(протокол № 12 от 20.02.2024)*

*Одобрено кафедрой «Математика и информатика»
(протокол № 13 от 27.02.2024)*

Краснодар 2024

УДК: 004:02
ББК: 32.973
Д95

Рецензенты: Е.А. Демехин, профессор кафедры «Математика и информатика» Краснодарского филиала Финансового университета при Правительстве РФ. Е.Н. Калайдин, доктор физико-математических наук, доцент, профессор кафедры «Математика и информатика» Краснодарского филиала Финуниверситета.

Дюдин М.С. Рабочая программа дисциплины технологии обработки данных для обучающихся по направлению 01.03.02 Прикладная математика и информатика, профиль «Анализ данных и принятие решений в экономике и финансах». – Краснодар: Краснодарский филиал Финуниверситета, кафедра «Математика и информатика», 2024 г.

Дисциплина Технологии обработки данных относится к Модулю общепрофессиональных дисциплин по направлению подготовки 01.03.02-Прикладная математика и информатика.

В рабочей программе дисциплины определены ее цель, требования к результатам освоения дисциплины, содержание программы, тематика аудиторных занятий, формы самостоятельной работы, оценочные средства для текущего контроля и промежуточной аттестации, учебно-методическое и информационное обеспечение.

Рабочая программа дисциплины технологии обработки данных

*Формат 60*90/16. Гарнитура Times New Roman*

Усл. п.л. 2,0. Изд. № _от.

Тираж 100 экз.

Заказ № .

Отпечатано в Краснодарском филиале Финуниверситета

© Дюдин М.С.
© Краснодарский филиал Финуниверситета, 2024

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	6
5.1. Содержание дисциплины.....	6
5.2. Учебно-тематический план	7
4.3. 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	13
7.3. Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний	21
7.4. Примерные вопросы для подготовки к экзамену	22
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	23
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	24
10. Методические указания для обучающихся по освоению дисциплины.....	24
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	25
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	25

1. Наименование дисциплины

Дисциплина «Технологии обработки данных»

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

Дисциплина «Технологии обработки данных» обеспечивает инструментарий формирования следующих компетенций: ПКН-2.

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКН-2	Способен с помощью математической модели решать поставленную теоретическую или прикладную задачу, реализовывая алгоритм решения в виде программного модуля	1. Демонстрирует знание базовых математических моделей, применяемых в различных предметных областях.	<p>Знать: положения теории вероятностей и мат. статистики, свойства непрерывных и дискретных случайных величин, типичные законы распределения</p> <p>Уметь: подбирать подходящий закон распределения для моделирования исследуемого показателя, определять его числовые характеристики используя алгоритмы на языке Python</p>
		2. Адаптирует и применяет существующие математические модели для решения поставленной прикладной или теоретической задачи.	<p>Знать: ограничения существующих теоретико-вероятностных моделей, методы мат. моделирования из смежных областей</p> <p>Уметь: проверять значимость построенной модели с использованием модулей Statistics, Pandas Python, выбирать оптимальный способ решения поставленной задачи</p>
		3. Владеет методологией математического моделирования для решения профессиональных задач.	<p>Знать: стандартные алгоритмы проверки статистических гипотез, составления вероятностных моделей особенности их реализации на Python</p> <p>Уметь: с помощью языка Python создавать мат. модели реальных экономических и IT процессов.</p>

3. Место дисциплины в структуре образовательной программы

Дисциплина «Технологии обработки данных» относится к Общепрофессиональному модулю дисциплин по направлению подготовки 01.03.02 -Прикладная информатика, профиль «Анализ данных и принятие решений в экономике и финансах».

Изучение дисциплины «Технологии обработки данных» основывается на сумме знаний, полученных при изучении дисциплины «Алгоритмы и структуры данных в языке Python». Для изучения данной дисциплины студент должен обладать базовыми знаниями в области информационных технологий и программирования, навыками программирования на языке Python.

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

Очная форма обучения

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Семестр 3 (в часах)
Общая трудоёмкость дисциплины	4/144	144
Контактная работа- Аудиторные занятия	50	50
Лекции	16	16
Семинары, практические занятия	34	34
Самостоятельная работа	94	94
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Экзамен	Экзамен

Очно – заочная форма обучения

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Семестр 4 (в часах)
Общая трудоёмкость дисциплины	4/144	144
Контактная работа- Аудиторные занятия	28	28
Лекции	12	12
Семинары, практические занятия	16	16
Самостоятельная работа	80	80
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Экзамен	Экзамен

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. Библиотека NumPy и Pandas.

В рамках темы рассматривается технологический стек Python для обработки и анализа данных, возможности Python как glue language, специфика библиотеки NumPy и ее роль в экосистеме Python. Организация массивов в NumPy: хранение данных, создание массивов, принципы реализации операций с едиными исходными данными. Универсальные функции и применение функций по осям в NumPy. Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры Маскирование и прихотливое индексирование в NumPy.

В рамках темы рассматриваются возможности библиотеки Pandas. Организация Pandas DataFrame и организация индексации для DataFrame и Series; применение универсальных функций и работа с пустыми значениями в Pandas. Объединение данных из нескольких Pandas DataFrame: общая логика и примеры. Рассматривается операция GroupBy в Pandas DataFrame и реализация в ней подхода «разбиение, применение и объединение».

Тема 2. Использование различных форматов файлов в задачах обработки данных.

В рамках темы рассматриваются принципы работы с файлами, файлы и операционные системы. Специфика текстовых и бинарных файлов.

В рамках темы рассматривается задача сериализации и десериализации данных и использование различных форматов файлов для ее решения. Описание формата файла JSON и пример описания данных в этом формате и взаимодействия с ним в Python.

В рамках темы рассматриваются формат XML и модель DOM: общая характеристика, пример описания данных в XML и DOM, работа с ними с помощью библиотеки BeautifulSoup.

В рамках темы рассматривается проблематика форматов файлов для хранения и обработки больших данных. Форматы файлов NPY и HDF: общая характеристика, пример взаимодействия с данными этих форматов в Python.

Тема 3. Взаимодействие с табличными данными в приложениях обработки данных.

В рамках темы рассматривается формат файлов CSV, представление данных в этом формате и взаимодействие с ним в Python.

В рамках темы рассматриваются возможности использования Excel для внешних приложений обработки данных. Взаимодействие с Excel из Python с помощью библиотеки XLWings: принципы работы и примеры использования.

Тема 4. Визуализация данных.

В рамках темы рассматриваются основы работы с библиотекой `matplotlib`: организация системы координат, оформление осей, цвета и цветовые карты в `matplotlib`, стили линий и маркеры. `Pyplot` и объектно-ориентированный интерфейс `matplotlib`. Управление фигурами и создание множества графиков на одном рисунке. Различные типы графиков.

В рамках темы рассматривается визуализация данных с помощью библиотеки `Pandas`: набор методов для построения графиков, реализованный в структурах `Series` и `DataFrame`.

В рамках темы проводится введение в разведочный анализ данных: типы признаков, анализ распределений, анализ мер центральной тенденции и поиск выбросов, анализ взаимного распределения и парных корреляций. Проведение разведочного анализа данных с помощью библиотеки `Seaborn`.

Тема 5. Работа со строками в приложениях обработки данных.

В рамках темы рассматриваются возможности `python` по форматированию строк: `%`-форматирование, метод `format`, `f`-строки.

В рамках темы рассматриваются основы работы с регулярными выражениями: базовый синтаксис, примеры. Модуль `re` в `Python`. Примеры использования регулярных выражений.

В рамках темы рассматривается использования хэширования при работе со строками. Строки в библиотеке `numru`.

Тема 6. Введение в обработку текста на естественном языке в задачах обработки данных.

В рамках темы рассматриваются сегментация и токенизация текста на естественном языке, стемминг и лемматизация, примеры на `Python`. Использование мемоизации на примере работы со строками. Расстояние Ле-венштейна: определение, алгоритм эффективного поиска оптимального редакционного предписания, пример поиска на `Python`. Векторное представление текста на естественном языке: общий алгоритм подходов `TF`; `TF-IDF`.

Тема 7. Профилирование процессов обработки данных, библиотека `Numba` и векторизация в `Numru` и `Numba`.

В рамках темы рассматривается профилирование реализации алгоритмов на `Python`, принципы решения задачи оптимизации производительности алгоритма. Библиотека `Numba`: принципы работы, базовые примеры использования. Векторизация в `numru`: ключевые параметры функции, примеры применения, использование обобщенной сигнатуры функции.

5.2. Учебно-тематический план

Очная форма обучения

		Трудоемкость в часах	
--	--	----------------------	--

№ п/п	Наименование темы (раздела) дисциплины	Всего	Контактная работа- Аудиторная работа			Самостоя- тельная работа	Формы текущего контроля успеваемости
			Общая, в т.ч.:	Лекции	Семинары, практические занятия		
1	Библиотека NumPy и Pandas	28	14	4	10	14	Самостояте- льные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
2	Использование различных форматов файлов в задачах обработки данных.	20	6	2	4	14	
3	Взаимодействие с табличными данными в приложениях обработки данных.	20	6	2	4	14	
4	Визуализация данных	20	6	2	4	14	
5	Работа со строками в приложениях обработки данных	20	6	2	4	14	
6	Введение в обработку текста на естественном языке в задачах обработки данных	20		2	4	14	
7	Профилирование процессов обработки данных, библиотека Numba и векторизация в Numpy и Numba	16		2	4	10	
В целом по дисциплине		144	50	16	34	94	Согласно учебному плану: контрольная работа

Очно – заочная форма обучения

№ п/п	Наименование темы (раздела) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа- Аудиторная работа			Самостоя- тельная работа	
			Общая, в т.ч.:	Лекции	Семинары, практические занятия		
1	Библиотека NumPy и Pandas	28	4	1	5	10	Самостояте- льные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям.
2	Использование различных форматов файлов в задачах обработки данных.	20	4	2	2	10	
3	Взаимодействие с табличными данными в приложениях обработки данных.	20	4	2	3	10	
4	Визуализация данных	20	4	1	1	10	
5	Работа со строками в приложениях обработки данных	20	4	3	3	20	
6	Введение в обработку текста на естественном языке в задачах обработки данных	20		1	1	10	
7	Профилирование процессов обработки данных, библиотека Numba и векторизация в Numpy и Numba	16		1	1	10	
В целом по дисциплине		144	28	12	16	80	Согласно учебному плану: контрольная работа

4.3.Содержание семинаров, практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8, 9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Библиотека NumPy и Pandas	<ul style="list-style-type: none"> • Технологический стек Python для обработки и анализа данных • Возможности Python как glue language • Организация массивов в NumPy: хранение данных, создание массивов • Принципы реализации операций с едиными исходными данными. Универсальные функции и применение функций по осям в NumPy. • Организация Pandas DataFrame и организация индексации для DataFrame и Series. • Применение универсальных функций и работа с пустыми значениями в Pandas. • Объединение данных из нескольких Pandas DataFrame: общая логика и примеры. <p>9[1], 8[2], 8[3]</p>	Интерактивная форма, работа на компьютере
Использование различных форматов файлов в задачах обработки данных	<ul style="list-style-type: none"> • Формат файлов Pickle, представление данных в этом формате и взаимодействие с ним в Python. • Формат файлов JSON, представление данных в этом формате и взаимодействие с ним в Python. • Формат XML и модель DOM: общая характеристика, пример описания данных в XML и DOM • Работа с XML с помощью библиотеки BeautifulSoup. <p>9[1], 8[2], 8[3]</p>	Интерактивная форма, работа на компьютере
Взаимодействие с табличными данными в приложениях обработки данных.	<ul style="list-style-type: none"> • Взаимодействие с Excel из Python с помощью библиотеки XLWings. • Формат файлов CSV, представление данных в этом формате и взаимодействие с ним в Python <p>9[1], 8[2], 8[3]</p>	Интерактивная форма, работа на компьютере
Визуализация данных	<ul style="list-style-type: none"> • Построение визуализаций с помощью библиотеки matplotlib <ul style="list-style-type: none"> • Построение визуализаций с помощью библиотеки pandas • Построение визуализаций с помощью библиотеки seaborn <p>9[1], 8[2], 8[3]</p>	Интерактивная форма, работа на компьютере
Работа со строками в приложениях обработки данных	<ul style="list-style-type: none"> • Основы работы с регулярными выражениями: базовый синтаксис, примеры. • Модуль re в Python. <p>9[1], 8[2], 8[3]</p>	Интерактивная форма, работа на компьютере

Введение в обработку текста на естественном языке в задачах обработки данных.	<ul style="list-style-type: none"> Сегментация и токенизация текста на естественном языке, стемминг и лемматизация, примеры на Python. Расстояние Левенштейна: определение, алгоритм эффективного поиска оптимального редакционного предписания, пример поиска на Python. 9[1], 8[2], 8[3]	Интерактивная форма, работа на компьютере
Профилирование процессов обработки данных, библиотека Numba и векторизация в Numpy и Numba	<ul style="list-style-type: none"> профилирование реализации алгоритмов на Python принципы решения задачи оптимизации производительности алгоритма Библиотека Numba: принципы работы, базовые примеры использования. 9[1], 8[2], 8[3]	Интерактивная форма, работа на компьютере

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

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

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Библиотека NumPy и Pandas	<p>Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры.</p> <p>Маскирование и прихотливое индексирование в NumPy.</p> <p>Операция GroupBy в Pandas DataFrame и реализация в ней подхода «разбиение, применение и объединение».</p>	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Использование различных форматов файлов в задачах обработки данных	<p>Формат файлов NPY, представление данных в этом формате и взаимодействие с ним в Python.</p> <p>Формат файлов HDF, представление данных в этом формате и взаимодействие с ним в Python.</p>	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Взаимодействие с табличными данными в приложениях обработки данных.	<ul style="list-style-type: none"> Продвинутые операции с Excel из Python с помощью библиотеки XLWings. 	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Визуализация данных	<ul style="list-style-type: none"> Построение трехмерных графиков Продвинутая работа с цветовыми картами 	Обзор литературы и веб-источников. Самостоятельное освоение

		инструментов аналитической обработки. Решение задач.
Работа со строками в приложениях обработки данных	Использования хэширования при работе со строками. Строки в библиотеке numpy.	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Введение в обработку текста на естественном языке в задачах обработки данных.	Использование мемоизации на примере работы со строками. Векторное представление текста на естественном языке: общий алгоритм подходов TF; TF-IDF.	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.
Профилирование процессов обработки данных, библиотека Numba и векторизация в Numpy и Numba	Векторизация в numpy: ключевые параметры функции, примеры применения Использование обобщенной сигнатуры функции в numpy и numba. •	Обзор литературы и веб-источников. Самостоятельное освоение инструментов аналитической обработки. Решение задач.

6.2.Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерные вопросы к контрольной работе

1. Организация массивов в NumPy: хранение данных, создание массивов, принципы реализации операций с едиными исходными данными
2. Универсальные функции и применение функций по осям в NumPy
3. Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры
4. Маскирование и прихотливое индексирование в NumPy
5. Векторизация в numpy: ключевые параметры функции, примеры применения, использование обобщенной сигнатуры функции

Примерные задания контрольной работы

1. В массиве чисел, хранящихся в файле finance.csv найти строку (вывести ее индекс и содержащиеся значения), в которой более всего значений, превышающих среднее значение по всему массиву. Для расчётов использовать Pandas.
2. В массиве чисел, хранящихся в файле finance.csv, подсчитать количество

строк, в которых более 600 значений больше среднего значения по всему массиву. Для расчётов использовать Pandas.

3. В массиве чисел, хранящихся в файле `finance.csv`, подсчитать количество значений, не отклоняющихся от среднего значения более чем на 3 стандартных отклонения. Для расчетов использовать Pandas.

Критерии балльной оценки различных форм текущего контроля успеваемости

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях кафедры «Математика и информатика».

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

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 1. «Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Технологии обработки данных».

Фонд оценочных средств включает контрольные материалы для проведения текущего контроля и промежуточной аттестации.

7. 1 Описание показателей и критериев оценивания компетенций

Планируемые результаты освоения компетенции (индикатора достижения компетенции)	Уровень освоения				Оценочное средство
	«неудовлетвори- тельно»	«удовлетворит ельно»	«хорошо»	«отлично»	
ПКН-2 Способен с помощью математической модели решать поставленную теоретическую или прикладную задачу, реализовывая алгоритм решения в виде программного модуля					
1. Демонстрирует знание базовых математических моделей, применяемых в различных предметных областях.					
Знать: —положения теории вероятностей и мат. статистики, свойства непрерывных и дискретных	Фрагментарное представление	Неполные представления	Сформированн ые, но содержащие отдельные пробелы представления	Сформированн ые систематически е представления	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания

Планируемые результаты освоения компетенции (индикатора достижения компетенции)	Уровень освоения				Оценочное средство
	неудовлетвори тельно»	«удовлетворит ельно»	«хорошо»	«отлично»	
случайных величин, типичные законы распределения					
Уметь: —подбирать подходящий закон распределения для моделирования исследуемого показателя, определять его числовые характеристики используя алгоритмы на языке Python	Фрагментарное умение	Несистематичес кое применение умений	В целом успешное, но содержащее отдельные пробелы умение	Сформированно е умение	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания
2. Адаптирует и применяет существующие математические модели для решения поставленной прикладной или теоретической задачи.					
Знать: —ограничения существующих теоретико- вероятностных моделей, методы мат. моделирования из смежных областей	Фрагментарное представление	Неполные представления	Сформированн ые, но содержащие отдельные пробелы представления	Сформированн ые систематически е представления	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания
Уметь: —проверять значимость построенной модели с использованием модулей Statistics, Pandas Python, выбирать оптимальный способ решения	Фрагментарное умение	Несистематичес кое применение умений	В целом успешное, но содержащее отдельные пробелы умение	Сформированно е умение	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания

Планируемые результаты освоения компетенции (индикатора достижения компетенции)	Уровень освоения				Оценочное средство
	неудовлетвори тельно»	«удовлетворит ельно»	«хорошо»	«отлично»	
поставленной задачи					
3. Владеет методологией математического моделирования для решения профессиональных задач.					
Знать: –стандартные алгоритмы проверки статистических гипотез, составления вероятностных моделей особенности их реализации на Python	Фрагментарное представление	Неполные представления	Сформированн ые, но содержащие отдельные пробелы представления	Сформированн ые систематически е представления	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания
Уметь: -с помощью языка Python создавать мат. модели реальных экономических и IT процессов.	Фрагментарное умение	Несистематичес кое применение умений	В целом успешное, но содержащее отдельные пробелы умение	Сформированно е умение	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания

7.2 Задания или иные материалы, необходимые для оценки знаний и умений, характеризующих формирование компетенций в процессе освоения ОП ВО

7.2.1 Вопросы для оценки знаний и умений, характеризующих формирование компетенций

Шифр компетенции	Вопросы	Правильный ответ
ПКН-2	1. Какой показатель надо вычислить, чтобы определить наиболее частое значение в наборе данных?	мода
	2. Напишите название модуля Python, который содержит статистические функции	statistics
	3. Напишите команду Python для вычисления медианы по списку x	statistics.median(x)
	4. Какой модуль Python позволяет загружать таблицы в структуру данных data frame?	pandas

5. Какой показатель оценивает средний квадрат отклонения от среднего?	дисперсия
6. Каким символом отделяются друг от друга значения в файлах .csv?	запятая
7. Какой тип переменной следует использовать в Python для действительных чисел?	float
8. Какой закон распределения описывается кривой Гаусса?	нормальный закон
9. От чего зависит достоверность оценки среднего значения показателя?	от объема выборки и дисперсии
10. Как меняется стандартная ошибка среднего при увеличении объема выборки?	уменьшается
11. Как называют значения, которые сильно выбиваются из набора данных, не вписываются в модель по какой-либо причине?	выбросы
12. Какие форматы файлов с данными может загружать программа, написанная на Python?	любые, если алгоритм может их обработать
13. Как называется библиотека Python, которая используется для визуализации данных, построения графиков, диаграмм	matplotlib
14. Какой критерий используется для проверки закона распределения?	критерий Пирсона или критерий Колмогорова
15. Как называется показатель, оценивающий взаимосвязь двух величин? Как его вычислить в Python?	коэффициент корреляции; <code>numpy.corrcoef(x,y)</code>
16. Какой модуль Python применяется для проверки стат. гипотез?	<code>scipy.stats</code>
17. Если надо проверить, что среднее больше определенного значения, то какой критерий надо применить?	критерий Стьюдента (t-статистика)
18. Какой метод применяется для определения коэффициентов линейной регрессии?	метод наименьших квадратов
19. Какие значения чаще всего принимают величины, подчиняющиеся нормальному закону распределения?	значения возле среднего
20. В чем смысл правила 3 сигма?	отклонение от среднего более чем на 3 стандартных отклонения маловероятно

7.2.2 Практико-ориентированные задания

Шифр компетенции	Практико-ориентированные задания	Правильный ответ
ПКН-2	1. Напишите программу, которая считывает данные в 4-м столбце таблицы в файле "GAZP.csv" в список	<pre>f=open('GAZP.csv','r') x=[] for i in range(10): z=f.readline().split(',')[3] try: z=float(z) x.append(z) except: pass</pre>

2. Напишите программу, загружающую таблицу из файла как data frame	<pre>import pandas as pd table1= pd.read_csv('GAZP.csv') df=pd.DataFrame(table1)</pre>
3. Напишите программу, оценивающую центр распределения данных, содержащихся в списке y.	<pre>import statistics mean_y=statistics.mean(y) median_y=statistics.median(y) try: mode_y=statistics.mode(y) except: mode_y='не существует'</pre>
4. Напишите программу, составляющую словарь, где ключи – названия показателей (среднее, ст отклонение), а значения – их величины.	<pre>import statistics mean_y=statistics.mean(y) stdev_y=statistics.stdev(y) dict_y=dict_y={'среднее':mean_y,'ст отклонение':stdev_y}</pre>
5. Как можно импортировать функцию mean() из библиотеки statistics, чтобы не указывать название библиотеки при её использовании?	<pre>from statistics import mean print(mean([1,2,3]))</pre>
6. Вычислите коэффициент вариации в Python	<pre>import statistics mean_y=statistics.mean(y) stdev_y=statistics.stdev(y) if mean_y != 0: c_var=stdev_y/mean_y else: c_var='не существует, т.к. среднее равно 0'</pre>
7. Посчитайте размах данных по списку y	<pre>min_y=min(y) max_y=max(y) range_y=max_y-min_y</pre>
8. Постройте график по 3 точкам, используя библиотеку matplotlib	<pre>import matplotlib.pyplot as plt x = [1,2,3] y = [2,4,1] plt.plot(x, y) plt.xlabel('x - axis') plt.ylabel('y - axis') plt.title('graph') plt.show()</pre>
9. Напишите программу для построения графика по загруженной из файла таблице	<pre>import matplotlib.pyplot as plt import pandas as pd df = pd.read_csv('fdata.csv', sep=',', parse_dates=True, index_col=0) df.plot() plt.show()</pre>
10. Посчитайте коэффициент корреляции Пирсона между двумя показателями	<pre>import numpy as np x = np.array([0, 1, 3]) y = np.array([2, 4, 5]) print(np.corrcoef(x, y))</pre>

	<p>11. Посчитайте дневные доходности акций «Газпром», их среднее и стандартное отклонение</p>	<pre>from statistics import mean, stdev f=open('GAZP.csv','r') x=[] for i in range(10): z=f.readline().split(',')[3] try: z=float(z) x.append(z) except: pass for i in range(len(x)-1): d[i]=(x[i]-x[i-1])/x[i-1] print(mean(d),stdev(d))</pre>
	<p>12. Найдите среди всех заработных плат (Wage) самую высокую и запишите это значение в переменную rich. Посчитайте, сколько зарабатывают сотрудники за год (Wage) вместе, и сохраните значение в переменной total. Выведите результаты кода через запятую: rich, low, total.</p>	<pre>import pandas as pd table1 = pd.read_csv('personel.csv') rich = table1.Wage.max() low = table1.Wage.min() total = table1.Wage.sum() print(rich, low, total)</pre>
	<p>13. Напишите программу, которая загрузит из файла таблицу с данными по сотрудникам и посчитает средний возраст, округленный до целых (столбец Age)</p>	<pre>import pandas as pd table1 = pd.read_csv('personel.csv') age = round(table1.Age.mean())</pre>
	<p>14. Напишите программу на Python для построения двух или более линий с легендами, различной ширины и цвета.</p>	<pre>import matplotlib.pyplot as plt # line 1 x1 = [10,20,30] y1 = [20,40,10] # line 2 x2 = [10,20,30] y2 = [40,10,30] plt.xlabel('x - axis') plt.ylabel('y - axis') plt.title('Two or more lines') plt.plot(x1,y1, color='blue', linewidth = 3, label = 'line1-width-3') plt.plot(x2,y2, color='red', linewidth = 5, label = 'line2-width-5') plt.legend() plt.show()</pre>
	<p>15. Напишите программу на Python для отображения сетки и нарисуйте линейные графики значения закрытия Alphabet Inc. в период с 3 октября 2016 года по 7 октября 2016 года. Настройте линии сетки с рендерингом с большей сеткой (основной сеткой) и меньшей</p>	<pre>import datetime as DT from matplotlib import pyplot as plt from matplotlib.dates import date2num data = [(DT.datetime.strptime('2016-10-03', "%Y-%m-%d"), 772.559998), (DT.datetime.strptime('2016-10-04', "%Y-%m-%d"), 776.429993), (DT.datetime.strptime('2016-10-05', "%Y-%m-%d"), 776.469971), (DT.datetime.strptime('2016-10-06', "%Y-%m-%d"), 776.859985), (DT.datetime.strptime('2016-10-07', "%Y-%m-%d"), 775.080017)]</pre>

	сетка (второстепенная сетка). Включите сетку, но отключите галочки.	<pre> x = [date2num(date) for (date, value) in data] y = [value for (date, value) in data] fig = plt.figure() graph = fig.add_subplot(111) # Plot the data as a red line with round markers graph.plot(x,y,'r-o') # Set the xtick locations graph.set_xticks(x) # Set the xtick labels graph.set_xticklabels([date.strftime("%Y-%m-%d") for (date, value) in data]) # naming the x axis plt.xlabel('Date') # naming the y axis plt.ylabel('Closing Value') # giving a title plt.title('Closing stock value of Alphabet Inc.') # Turn on the minor TICKS, which are required for the minor GRID plt.minorticks_on() # Customize the major grid plt.grid(which='major', linestyle='-', linewidth='0.5', color='red') # Customize the minor grid plt.grid(which='minor', linestyle=':', linewidth='0.5', color='black') # Turn off the display of all ticks. plt.tick_params(which='both', # Options for both major and minor ticks top='off', # turn off top ticks left='off', # turn off left ticks right='off', # turn off right ticks bottom='off') # turn off bottom ticks plt.show() </pre>
	16. Напишите программу на Python для отображения гистограммы популярности языков программирования.	<pre> import matplotlib.pyplot as plt x = ['Java', 'Python', 'PHP', 'JavaScript', 'C#', 'C++'] popularity = [22.2, 17.6, 8.8, 8, 7.7, 6.7] x_pos = [i for i, _ in enumerate(x)] plt.bar(x_pos, popularity, color='blue') plt.xlabel("Languages") plt.ylabel("Popularity") plt.title("Popularity of Programming Language\n" + "Worldwide, Oct 2017 compared to a year ago") plt.xticks(x_pos, x) # Turn on the grid plt.minorticks_on() plt.grid(which='major', linestyle='-', linewidth='0.5', color='red') # Customize the minor grid plt.grid(which='minor', linestyle=':', linewidth='0.5', color='black') plt.show() </pre>
	17. Напишите программу с использованием модуля Pandas для удаления последних n строк таблицы, записанной в структуру данных DataFrame.	<pre> import pandas as pd d = {'col1': [1, 2, 3, 4, 7, 11], 'col2': [4, 5, 6, 9, 5, 0], 'col3': [7, 5, 8, 12, 1, 11]} df = pd.DataFrame(data=d) print("Original DataFrame") print(df) print("\nAfter removing last 3 rows of the said DataFrame:") df1 = df.iloc[:3] print(df1) </pre>
	18. Решите задачу на построение линейной регрессии по двум спискам числовых значений	<pre> import numpy as np from sklearn.linear_model import LinearRegression x = np.array([5, 15, 25, 35, 45, 55]).reshape((-1, 1)) y = np.array([5, 20, 14, 32, 22, 38]) model = LinearRegression() model.fit(x, y) model = LinearRegression().fit(x, y) y_pred = model.predict(x) print('predicted response:', y_pred, sep='\n') y_pred = model.intercept_ + model.coef_ * x print('predicted response:', y_pred, sep='\n') </pre>

7.2.3 Тесты

Шифр компетенции	Тестовые задания	Правильный ответ
ПКН-2	1. Число, делящее множество чисел пополам, так, что 50% значений больше него, 50% меньше - 1) медиана 2) мода 3) дисперсия 4) среднее	1
	2. Для оценки разброса данных вычисляют: 1) дисперсию 2) моду 3) ассиметрию 4) эксцесс	1
	3. Чтобы получить доступ к 1-му элементу списка x надо написать: 1) x[1] 2) x[0] 3) x(1) 4) x(0)	2
	4. Функция mode содержится в библиотеке 1) statistics 2) numpy 3) pandas 4) scipy	1
	5. Если медиана и мода больше среднего, можно ожидать, что: 1) чаще будут встречаться значения выше среднего 2) чаще будут встречаться значения ниже среднего 3) значения выше и ниже среднего будут встречаться с равной вероятностью	1
	6. Программа, написанная на Python, превосходит расчеты в табличном процессоре 1) скоростью вычислений 2) простотой создания 3) объемом на жестком диске 4) объемом в оперативной памяти	1
	7. Функции и методы для построения графиков содержатся в библиотеке: 1) statistics 2) matplotlib 3) numpy 4) pandas	2
	8. Функция stdev содержится в библиотеке 1) statistics 2) numpy 3) pandas 4) scipy	1
	9. Какие команды можно использовать для загрузки столбца таблицы, содержащего текстовые и числовые значения вперемешку?	1

	1) try except 2) if else 3) float, str	
	10. Какой закон распределения лучше всего описывает распределение доходностей финансового актива? 1) нормальный 2) показательный 3) равномерный 4) гама	1
	11. Для моделирования временного интервала между событиями целесообразнее всего применять: 1) нормальный закон распределения 2) показательный закон распределения 3) равномерный закон распределения 4) гама распределение	2
	12. При проверке стат. гипотезы о числовом значении показателя нулевая гипотеза обычно формулируется о 1) равенстве 2) неравенстве	1

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

Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
ПКН-2 Способен с помощью математической модели решать поставленную теоретическую или прикладную задачу, реализовывая алгоритм решения в виде программного модуля	1. Демонстрирует знание базовых математических моделей, применяемых в различных предметных областях.	<u>Знать:</u> положения теории вероятностей и мат. статистики, свойства непрерывных и дискретных случайных величин, типичные законы распределения <u>Уметь:</u> подбирать подходящий закон распределения для моделирования исследуемого показателя, определять его числовые характеристики используя алгоритмы на языке Python	Задание 1 Перечислите показатели, оценивающие центр распределения данных, команды Python для их вычисления. Задание 2 Загрузите столбец с ценами на открытие из файла "SBER.csv", вычислите темпы прироста, вычислите среднее и стандартное отклонение, подберите закон распределения, наилучшим образом подходящий к данным
	2. Адаптирует и применяет	<u>Знать:</u> ограничения существующих	Задание 1

	<p>существующие математические модели для решения поставленной прикладной или теоретической задачи.</p>	<p>теоретико-вероятностных моделей, методы мат. моделирования из смежных областей</p> <p>Уметь: проверять значимость построенной модели с использованием модулей Statistics, Pandas Python, выбирать оптимальный способ решения поставленной задачи</p>	<p>Среднее значение экономического показателя вычисляется по выборке из 16 значений, стандартное отклонение равно 10. Можно ли ожидать, что оценка среднего будет содержать ошибку более 15?</p> <p>Задание 2 Вычислите доверительный интервал генеральной доли отрицательных значений по выборке значений из файла "data1.txt"</p> <p>Задание 3 В файле "data2.csv" содержатся значения переменных x и y. Проведите корреляционно-регрессионный анализ</p>
	<p>3. Владеет методологией математического моделирования для решения профессиональных задач.</p>	<p>Знать: стандартные алгоритмы проверки статистических гипотез, составления вероятностных моделей особенности их реализации на Python</p> <p>Уметь: с помощью языка Python создавать мат. модели реальных экономических и IT процессов.</p>	<p>Задание 1 В файле "data1.csv" содержатся значения доходностей финансового актива. Проверьте гипотезу о том, что среднее доходностей выше 1%.</p> <p>Задание 2 Загрузите данные о динамике экономического показателя из файла "data2.csv", с помощью линейной регрессии спрогнозируйте его значения на 10 дней вперед.</p>

7.4.Примерные вопросы для подготовки к экзамену

1. Организация массивов в NumPy: хранение данных, создание массивов, принципы реализации операций с едиными исходными данными
2. Универсальные функции и применение функций по осям в NumPy
3. Принцип распространения значений при выполнении операций в NumPy: общий алгоритм и примеры
4. Маскирование и прихотливое индексирование в NumPy
5. Векторизация в numpy: ключевые параметры функции, примеры применения, использование обобщенной сигнатуры функции

6. Numba: принципы работы, базовые примеры использования
7. Организация Pandas DataFrame и организация индексации для DataFrame и Series
8. Применение универсальных функций и работа с пустыми значениями в Pandas
9. Объединение данных из нескольких Pandas DataFrame: общая логика и примеры
10. Операция GroupBy в Pandas DataFrame и реализация в ней подхода «разбиение, применение и объединение»
11. Специфика текстовых и бинарных файлов, форматы файлов CSV и Pickle, представление данных в этих форматах и взаимодействие с ними в Python
12. Задача сериализации и десериализации, описание формата файла JSON и пример описания данных в этом формате и взаимодействия с ним в Python
13. Формат XML и модель DOM: общая характеристика, пример описания данных в XML и DOM, работа с ними с помощью библиотеки BeautifulSoup
14. Форматы файлов NPY и HDF общая характеристика, пример взаимодействия с данными этих форматов в Python
15. Взаимодействие из Python с базой данных на примере API SQLite, базовые возможности работы с транзакциями
16. Взаимодействие с Excel из Python с помощью XLWings: принципы работы и примеры использования
17. Основы работы с регулярными выражениями: базовый синтаксис, примеры использования модуля re в Python
18. Сегментация и токенизация текста на естественном языке, стемминг и лемматизация, примеры на Python
19. Расстояние Левенштейна: определение, алгоритм эффективного поиска оптимального редакционного предписания, пример поиска на Python
20. Векторное представление текста на естественном языке: общий алгоритм подходов TF; TF-IDF.

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

Основная литература:

1. Колдаев, В. Д. Структуры и алгоритмы обработки данных : учебное пособие / В. Д. Колдаев. - Москва : РИОР : ИНФРА-М, 2021. - 296 с. - ЭБС ZNANIUM.com. - URL: <https://znanium.com/catalog/product/1230215> (дата обращения: 17.01.2023). - Текст: электронный.
2. Дадян, Э. Г. Методы, модели, средства хранения и обработки данных : учебник / Э.Г. Дадян, Ю.А. Зеленков. — Москва : Вузовский учебник : ИНФРА-М, 2024. — 168 с. - ISBN 978-5-9558-0490-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2122966>

Дополнительная литература:

3. Нагаева, И. А. Основы алгоритмизации и программирования: практикум : учебное пособие / И. А. Нагаева, И. А. Кузнецов. - Москва : Берлин : Директ-Медиа,

2021. - 169 с. - ЭБС Университетская библиотека ONLINE. - URL: <https://biblioclub.ru/index.php?page=book&id=598404> (дата обращения: 17.01.2023). - Текст : электронный.

4. Курзаева, Л. В. Основы обработки данных социологических исследований в электронных таблицах : учебное пособие / Л. В. Курзаева, Т. В. Макарова. - 2-е изд., стер. - Москва : ФЛИНТА, 2023. - 105 с. - ISBN 978-5-9765-3740-8. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2091316>

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

1. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
2. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
3. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>
4. Электронно-библиотечная система Znanium <http://www.znanium.com> Pylru 1.0.9 [Электронный ресурс]: сайт. – Режим доступа: <https://pypi.python.org/pypi/pylru>
5. Python Data Analysis Library [Электронный ресурс]: сайт. – Режим доступа: <http://pandas.pydata.org/>
6. Python Documentation [Электронный ресурс]: сайт. – Режим до- ступа: <http://python.org/doc/>
7. Python Standard Library [Электронный ресурс]: сайт. – Режим до- ступа: <https://docs.python.org/2/library/>
8. Scikit-learn Machine Learning in Python [Электронный ресурс]: сайт. – Режим доступа: <http://scikit-learn.org>
9. Официальный сайт продукта <https://www.python.org/>
10. Каталог курсов Интернет Университета Информационных Техно- логий <http://www.intuit.ru/>
11. The Python Tutorial // <https://docs.python.org/3/tutorial/index.html>
12. NumPy User Guide // <http://docs.scipy.org/doc/numpy/user/index.html>
13. Pandas User Guide <http://pandas.pydata.org/pandas-docs/stable/>
14. Dask User Guide <https://docs.dask.org/en/latest/>
15. Dask User Guide <https://docs.dask.org/en/latest/>
16. Matplotlib User Guide // <https://matplotlib.org/stable/users/index.html>
Seaborn User Guide // <https://seaborn.pydata.org/tutorial.html>

10.Методические указания для обучающихся по освоению дисциплины

При изучении теоретического материала необходимо опираться на рабочую программу дисциплины, материалы лекций и литературу из основного списка. Кроме этого, необходимо активно работать с Интернет-источниками и пособиями других авторов, помогающими усвоить материал отдельных разделов программы.

Необходимо конспектировать лекции, помечая сложные и непонятные моменты с тем, чтобы задать вопросы лектору в конце лекции или же на

консультации.

При подготовке к семинарским занятиям необходимо изучить вопросы, вынесенные на самостоятельное изучение, так как семинарские занятия предполагают их обсуждение и дискуссию по теме; кроме того, задания для самостоятельной работы необходимы для того, чтобы успешно выполнить самостоятельные задания на семинарах.

Индивидуальные задания для работы на компьютере, файлы с выполненными заданиями необходимо хранить в личной сетевой папке в компьютерной сети вуза.

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

1. Комплект безлицензионного программного обеспечения

1. Astra Linux
2. Python 3
3. PyCharm

2. Современные профессиональные базы данных и информационные справочные системы

не требуется

3. Сертифицированные программные и аппаратные средства защиты информации

- не используются.

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

Для проведения лекций и практических занятий необходима аудитория, оснащенная проектором и компьютерами с постоянным подключением к сети Интернет.