

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Макаренко Елена Николаевна

Должность: Ректор

Дата подписания: 10.12.2024 15:07:32

Уникальный программный ключ:

c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования «Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ

Начальник

учебно-методического управления

Платонова Т.К.

«25» июня 2024 г.

**Рабочая программа дисциплины
Статистический анализ данных в Python**

Направление 01.03.05 Статистика

Направленность 01.03.05.01 Анализ больших данных

Для набора 2024 года

Квалификация
Бакалавр

КАФЕДРА Информационных систем и прикладной информатики**Распределение часов дисциплины по семестрам**

Семестр (<Курс>.<Семестр на курсе>)	4 (2.2)		Итого	
	16			
Неделя	16			
Вид занятий	УП	РП	УП	РП
Лабораторные	64	64	64	64
Итого ауд.	64	64	64	64
Контактная работа	64	64	64	64
Сам. работа	116	116	116	116
Часы на контроль	36	36	36	36
Итого	216	216	216	216

ОСНОВАНИЕ

Учебный план утвержден учёным советом вуза от 25.06.2024 г. протокол № 18.

Программу составил(и): д.э.н., доц., Щербаков С.М.; ст. преп., Шкодина Т.А.

Зав. кафедрой: д.э.н., проф. Щербаков С.М.

Методический совет направления: к.э.н., доцент Андреева О.В.

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	освоение базовых основ статистической методологии сбора и обработки данных, представления полученной на их основе информации в виде статистических моделей и формирования выводов, направленных на решение практических задач с помощью высокоуровневого язык программирования python.
-----	--

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ПК-6: Способен осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием технологий больших данных

ПК-1: Способен разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи

В результате освоения дисциплины обучающийся должен:

Знать:

основные структуры представления и хранения данных , а также алгоритмы, используемые для их обработки (соотнесено с индикатором ПК-1.1)

основные методы и технологии первичной обработки данных для осуществления поиска информации в больших данных (соотнесено с индикатором ПК-6.1)

Уметь:

разрабатывать алгоритмы для решения поставленных задач в виде программного модуля с использованием структур данных (соотнесено с индикатором ПК -1.2)

обрабатывать большие данные на языке программирования python(соотнесено с индикатором ПК-6.2)

Владеть:

навыками программной реализации алгоритмов на языке программирования python (соотнесено с индикатором ПК - 1.3)

навыками программной реализации методов и моделей на языке программирования python(соотнесено с индикатором ПК-6.3)

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Раздел 1. Статистические библиотеки в python

№	Наименование темы / Вид занятия	Семестр / Курс	Часов	Компетенции	Литература
1.1	Математический аппарат языка Python. Введение в модуль Numpy / Лаб /	4	4	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2
1.2	Статистика и комбинаторика. Математическая библиотека Python – SymPy / Лаб /	4	4	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2

Раздел 2. Статистический анализ данных

№	Наименование темы / Вид занятия	Семестр / Курс	Часов	Компетенции	Литература
2.1	Анализ данных. Основы работы с Pandas / Лаб /	4	4	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2
2.2	Работа с текстовыми данными. Сбор данных из открытых источников. Предварительная обработка текстовых данных / Лаб /	4	8	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2
2.3	Построение и анализ сетей графов. Библиотека Networkx / Лаб /	4	4	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2
2.4	Модели и алгоритмы машинного обучения. Оценка алгоритмов машинного обучения. Прогнозирование данных. Библиотека Scipy, Sklearn / Лаб /	4	6	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2
2.5	Иерархическая кластеризация. Деревья решений. Случайный лес. Ансамбли моделей. Модуль Sklearn / Лаб /	4	6	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2
2.6	Обучение без учителя. Алгоритм кластеризации. Метод k-средних / Лаб /	4	8	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2
2.7	Визуализация данных. Представление результатов исследования. Построение графического изображения результатов статистического анализа данных / Лаб /	4	6	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2

2.8	Обработка и визуализация графических изображений / Лаб /	4	8	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2
2.9	Создание нейронной сети. Использование библиотеки Keras / Лаб /	4	6	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2
2.10	Выполнение индивидуального задания. Темы заданий представлены в приложении 1 к рабочей программе дисциплины. / Ср /	4	116	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2
2.11	Экзамен / Экзамен /	4	36	ПК-6, ПК-1	Л1.1, Л1.2, Л2.1, Л2.2

4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Структура и содержание фонда оценочных средств для проведения текущей и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Основная литература

	Авторы,	Заглавие	Издательство, год	Колич-во
Л1.1	Лемешко, Б. Ю., Лемешко, С. Б., Постовалов, С. Н., Чимитова, Е. В.	Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: монография	Новосибирск: Новосибирский государственный технический университет, 2011	https://www.iprbookshop.ru/47719.html неограниченный доступ для зарегистрированных пользователей
Л1.2	Каган Е. С.	Прикладной статистический анализ данных: учебное пособие	Кемерово: Кемеровский государственный университет, 2018	https://biblioclub.ru/index.php?page=book&id=573550 неограниченный доступ для зарегистрированных пользователей

5.2. Дополнительная литература

	Авторы,	Заглавие	Издательство, год	Колич-во
Л2.1		Прикладная информатика: журнал	Москва: Университет Синергия, 2022	https://biblioclub.ru/index.php?page=book&id=688854 неограниченный доступ для зарегистрированных пользователей
Л2.2	Карякин, М. И., Ватульян, К. А., Мнухин, Р. М.	Технологии программирования и компьютерный практикум на языке Python: учебное пособие	Ростов-на-Дону, Таганрог: Издательство Южного федерального университета, 2022	https://www.iprbookshop.ru/125718.html неограниченный доступ для зарегистрированных пользователей

5.3 Профессиональные базы данных и информационные справочные системы

Национальная электронная библиотека (НЭБ) - <https://rusneb.ru/>
ИСС «КонсультантПлюс»
ИСС «Гарант» <http://www.internet.garant.ru/>

5.4. Перечень программного обеспечения

Операционная система РЕД ОС
LibreOffice
Python IDLE
Pycharm Community

5.5. Учебно-методические материалы для студентов с ограниченными возможностями здоровья

При необходимости по заявлению обучающегося с ограниченными возможностями здоровья учебно-методические материалы предоставляются в формах, адаптированных к ограничениям здоровья и восприятия информации. Для лиц с нарушениями зрения: в форме аудиофайла; в печатной форме увеличенным шрифтом. Для лиц с нарушениями слуха: в форме электронного документа; в печатной форме. Для лиц с нарушениями опорно-двигательного аппарата: в форме электронного документа; в печатной форме.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Помещения для всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения:

- столы, стулья;
- персональный компьютер / ноутбук (переносной);
- проектор;
- экран / интерактивная доска.

Лабораторные занятия проводятся в компьютерных классах, рабочие места в которых оборудованы необходимыми лицензионными программными средствами и выходом в Интернет.

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Методические указания по освоению дисциплины представлены в Приложении 2 к рабочей программе дисциплины.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

1.1. Показатели и критерии оценивания компетенций:

ЗУН, составляющие компетенцию	Показатели оценивания	Критерии оценивания	Средства оценивания
ПК-1: Способен разрабатывать и реализовывать в виде программного модуля алгоритм решения поставленной теоретической или прикладной задачи			
З. основные структуры представления и хранения данных, а также алгоритмы, используемые для их обработки	знает основные классы методов и алгоритмов для обработки структур данных.	полнота и содержательность ответа умение приводить примеры	Э – вопросы к экзамену (1-30)
У. разрабатывать алгоритмы для решения поставленных задач в виде программного модуля с использованием структур данных	ставит задачи, использует методы и алгоритмы для обработки различных структур данных	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ЛЗ – лабораторные задания (1-11), ИЗ-индивидуальные задания (1-5)
В. навыками программной реализации алгоритмов на языке программирования python	выполняет программную реализацию алгоритмов на языке python	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ЛЗ – лабораторные задания (1-11), ИЗ-индивидуальные задания (1-5)
ПК-6: Способен осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием технологий больших данных			
З. основные методы и технологии первичной обработки данных для осуществления поиска информации в больших данных	знает основные методы и технологии обработки данных, осуществляет поиск информации	полнота и содержательность ответа умение приводить примеры	Э – вопросы к экзамену (1-30)
У. обрабатывать большие данные на языке программирования python	осуществляет обработку больших данных	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ЛЗ – лабораторные задания (1-11), ИЗ-индивидуальные задания (1-5)
В. навыками программной реализации методов и моделей на языке программирования python	создает программное обеспечение с помощью языка программирования python на основе моделей и методов статистического анализа данных	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ЛЗ – лабораторные задания (1-11), ИЗ-индивидуальные задания (1-5)

1.2 Шкалы оценивания:

Текущий контроль успеваемости и промежуточная аттестация осуществляется в рамках накопительной балльно-рейтинговой системы в 100-балльной шкале:

84-100 баллов (оценка «отлично»);

67-83 баллов (оценка «хорошо»);

50-66 баллов (оценка «удовлетворительно»);

0-49 баллов (оценка «неудовлетворительно»).

2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Вопросы к экзамену

1. Синтаксис и основные конструкции языка программирования python
2. Встроенные операции и функции
3. Условный оператор. Множественное ветвление
4. Циклы и счетчики
5. Определение функций. Параметры и аргументы. Вызовы функций.
6. Списки и кортежи. Срезы
7. Множества и словари. Методы словарей
8. Случайные числа random, randrange
9. Функции обработки строк join, replace, split.
10. Стандартная библиотека и pip. Модули и пакеты в python
11. Файлы. Работа с внешними источниками
12. Работа с текстовыми файлами xml и csv
13. Генераторы, декораторы
14. ООП в python. Классы и объекты
15. Магические методы
16. Основные принципы организации распределенных систем обработки информации
17. Наука о данных и python. Библиотеки NumPy, pandas, scipy, matplotlib
18. Математический аппарат языка python. Модуль NumPy: массивы, векторные вычисления
19. Математическая библиотека python SymPy.
20. Инструменты визуализации данных. Библиотека matplotlib
21. Библиотека pandas. Введение в структуры данных pandas
22. Объекты Dataframe и Series
23. Визуализация данных в pandas. Seaborn
24. Научные расчеты с помощью библиотеки scipy. Модели и алгоритмы машинного обучения
25. Предварительная обработка текстовых данных. Работа с пропусками Веб-скрейпинг
26. Машинное обучение без учителя. Масштабирование данных
27. Задачи классификации и линейные модели
28. Машинное обучение без учителя. Кластеризация методом k-средних. Модуль Sklearn
29. Алгоритмы машинного обучения. Иерархическая кластеризация. Деревья решений. Случайный лес. Модуль Sklearn
30. Нейронные сети. Библиотека Keras

Экзаменационное задание включает три вопроса – два теоретических вопроса и одно практико-ориентированное задание из числа приведенных ниже лабораторных заданий.

Критерии оценивания:

- 84-100 баллов (оценка «отлично») – изложенный материал фактически верен, наличие глубоких исчерпывающих знаний в объеме пройденной программы дисциплины в соответствии с поставленными программой курса целями и задачами обучения; правильные, уверенные действия по применению полученных знаний на практике, грамотное и логически стройное изложение материала при ответе, усвоение основной и знакомство с дополнительной литературой;
- 67-83 баллов (оценка «хорошо») – наличие твердых и достаточно полных знаний в объеме пройденной программы дисциплины в соответствии с целями обучения, правильные действия по

применению знаний на практике, четкое изложение материала, допускаются отдельные логические и стилистические погрешности, обучающийся усвоил основную литературу, рекомендованную в рабочей программе дисциплины;

- 50-66 баллов (оценка «удовлетворительно») – наличие твердых знаний в объеме пройденного курса в соответствии с целями обучения, изложение ответов с отдельными ошибками, уверенно исправленными после дополнительных вопросов; правильные в целом действия по применению знаний на практике;

- 0-49 баллов (оценка «неудовлетворительно») – ответы не связаны с вопросами, наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания на практике, неуверенность и неточность ответов на дополнительные и наводящие вопросы.

Лабораторные задания

Лабораторное задание №1

Математический аппарат языка Python. Введение в модуль Numpy.

Лабораторное задание №2

Статистика и комбинаторика. Математическая библиотека Python – SymPy.

Лабораторное задание №3

Анализ данных. Основы работы с Pandas.

Лабораторное задание №4

Работа с текстовыми данными. Сбор данных из открытых источников. Предварительная обработка текстовых данных.

Лабораторное задание №5

Построение и анализ сетей графов. Библиотека Networkx.

Лабораторное задание №6

Модели и алгоритмы машинного обучения. Оценка алгоритмов машинного обучения. Прогнозирование данных. Библиотека Scipy, Sklearn.

Лабораторное задание №7

Иерархическая кластеризация. Деревья решений. Случайный лес. Ансамбли моделей. Модуль Sklearn.

Лабораторное задание №8

Обучение без учителя. Алгоритм кластеризации. Метод k-средних.

Лабораторное задание №9

Визуализация данных. Представление результатов исследования. Построение графического изображения результатов статистического анализа данных.

Лабораторное задание №10

Обработка и визуализация графических изображений.

Лабораторное задание №11

Создание нейронной сети. Использование библиотеки Keras.

Критерии оценивания (для каждого задания):

6-7 б. – задание выполнено верно;

4-5 б. – при выполнении задания были допущены неточности, не влияющие на результат;

2-3 б. – при выполнении задания были допущены ошибки;

0-1 б. – при выполнении задания были допущены существенные ошибки.

Максимальное количество баллов за практические задания – 77 (11 заданий по 7 баллов).

Индивидуальные задания

Индивидуальное задание 1. Используя сайт <https://www.kaggle.com/datasets> выбрать датасет по определенной тематике. Загрузить данный датасет в Google Colab. Постройте модель на основе данных, используя иерархическую кластеризацию и модуль sklearn.

Индивидуальное задание 2. Используя сайт <https://www.kaggle.com/datasets> выбрать датасет по определенной тематике. Загрузить данный датасет в Google Colab. Постройте модель на основе данных, используя метод k-средних и модуль sklearn.

Индивидуальное задание 3. Используя сайт <https://www.kaggle.com/datasets> выбрать датасет по определенной тематике. Загрузить данный датасет в Google Colab. Выполнить статистический анализ данных, используя модуль pandas.

Индивидуальное задание 4. Используя сайт <https://www.kaggle.com/datasets> выбрать датасет по определенной тематике. Загрузить данный датасет в Google Colab. Постройте модель на основе данных, используя иерархическую кластеризацию и модуль sklearn.

Индивидуальное задание 5. Используя сайт <https://www.kaggle.com/datasets> выбрать датасет по определенной тематике. Загрузить данный датасет в Google Colab. Постройте модель на основе данных, используя метод k-средних и модуль sklearn.

Критерии оценивания (для каждого задания):

20-23 б. – задание выполнено верно;

15-19 б. – при выполнении задания были допущены неточности, не влияющие на результат;

10-14 б. – при выполнении задания были допущены ошибки;

0-9 б. – при выполнении задания были допущены существенные ошибки.

Максимальное количество баллов за индивидуальные задания – 23

3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

Текущий контроль успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения студентов до промежуточной аттестации.

Промежуточная аттестация проводится в форме экзамена.

Экзамен проводится по расписанию промежуточной аттестации. Количество вопросов в задании – 3 (два теоретических вопроса и одно практико-ориентированное задание). Объявление результатов производится в день экзамена. Результаты аттестации заносятся в ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику, должны ликвидировать задолженность в установленном порядке.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Учебным планом предусмотрены следующие виды занятий:

- лабораторные занятия.

В ходе лабораторных занятий углубляются и закрепляются знания студентов по ряду рассмотренных на лекциях вопросов, развиваются навыки практической работы.

При подготовке к лабораторным занятиям каждый студент должен:

- изучить рекомендованную учебную литературу;
- подготовить ответы на все вопросы по изучаемой теме.

В процессе подготовки к лабораторным занятиям студенты могут воспользоваться консультациями преподавателя.

Вопросы, не рассмотренные на лабораторных занятиях, должны быть изучены студентами в ходе самостоятельной работы. Контроль самостоятельной работы студентов над учебной программой курса осуществляется в ходе занятий методом выполнения лабораторных и индивидуальных заданий. В ходе самостоятельной работы каждый студент обязан прочитать основную и по возможности дополнительную литературу по изучаемой теме, выделить непонятные термины, найти их значение в энциклопедических словарях.

Студент должен готовиться к предстоящему лабораторному занятию по всем обозначенным в рабочей программе дисциплины вопросам.

Для подготовки к занятиям, текущему контролю и промежуточной аттестации студенты могут воспользоваться электронно-библиотечными системами. Также обучающиеся могут взять на дом необходимую литературу на абонементе университетской библиотеки или воспользоваться читальными залами.