

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Галунин Сергей Александрович
Должность: проректор по учебной работе
Дата подписания: 26.11.2024 14:26:37
Уникальный программный ключ:
08ef34338325bdb0ac5a47baa5472ce36cc3fc3b

Приложение к ОПОП
«Организация и программирова-
ние интеллектуальных систем»



СПбГЭТУ «ЛЭТИ»
ПЕРВЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ

МИНОБРНАУКИ РОССИИ

федеральное государственное автономное образовательное учреждение высшего образования
**«Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И.Ульянова (Ленина)»
(СПбГЭТУ «ЛЭТИ»)**

РАБОЧАЯ ПРОГРАММА

дисциплины

«АНАЛИЗ ДАННЫХ В ПРОГРАММАХ С ОТКРЫтыМ КОДОМ»

для подготовки бакалавров

по направлению

09.03.01 «Информатика и вычислительная техника»

по профилю

«Организация и программирование интеллектуальных систем»

Санкт-Петербург

2024

ЛИСТ СОГЛАСОВАНИЯ

Разработчики:

доцент, к.т.н., доцент Шумилов Л.А.

Рабочая программа рассмотрена и одобрена на заседании кафедры ВТ
19.01.2024, протокол № 1

Рабочая программа рассмотрена и одобрена учебно-методической комиссией
ФКТИ, 24.01.2024, протокол № 1

Согласовано в ИС ИОТ

Начальник ОМОЛА Загороднюк О.В.

1 СТРУКТУРА ДИСЦИПЛИНЫ

Обеспечивающий факультет	ФКТИ
Обеспечивающая кафедра	ВТ
Общая трудоемкость (ЗЕТ)	3
Курс	3
Семестр	5

Виды занятий

Лекции (академ. часов)	34
Практические занятия (академ. часов)	34
Иная контактная работа (академ. часов)	1
Все контактные часы (академ. часов)	69
Самостоятельная работа, включая часы на контроль (академ. часов)	39
Всего (академ. часов)	108

Вид промежуточной аттестации

Зачет (курс) 3

2 АННОТАЦИЯ ДИСЦИПЛИНЫ

«АНАЛИЗ ДАННЫХ В ПРОГРАММАХ С ОТКРЫТЫМ КОДОМ»

В основу курса положена методическая концепция практикума по изучению и освоению инструментов для анализа данных. В качестве инструментов используются программные продукты с открытым кодом. Все практические работы практикума сгруппированы в разделы анализа данных, теоретическое обеспечение которых выполняется на лекционных занятиях

SUBJECT SUMMARY

«DATA ANALYSIS TOOLS»

The course is based on the methodological concept of the workshop on the study and development of tools for data analysis. The tools used are open source software products. All practical work of the workshop is grouped into sections of data analysis, the theoretical support of which is carried out in lectures

3 ОБЩИЕ ПОЛОЖЕНИЯ

3.1 Цели и задачи дисциплины

1. Целью дисциплины является:

- формирование базовых компетенций, позволяющих выработать обоснованные стратегические решения, сочетающие интуицию специалиста с тщательным анализом имеющейся информации;
- приобретение умений выбирать инструментальные средства для обработки многомерных данных в соответствии с поставленной задачей;
- формирование навыков выполнять выбор и применять инструментальные средства для обработки многомерных данных.

2. Основными задачами дисциплины являются: применение статистического инструментария в исследовании многомерных совокупностей, а также приобретение навыков в применении в профессиональной деятельности основных методов обработки и анализа данных, а также знания по методам многомерного статистического анализа данных, необходимые для решения поставленных задач и умение выбирать инструментальные средства для обработки многомерных данных в соответствии с поставленной задачей.

3. Знания методов многомерного статистического анализа данных, необходимых для решения поставленных задач.

4. Умения выбирать инструментальные средства для обработки многомерных данных в соответствии с поставленной задачей.

5. Владение навыками выбора и применения инструментальных средств для обработки многомерных данных.

3.2 Место дисциплины в структуре ОПОП

Дисциплина изучается на основе ранее освоенных дисциплин учебного плана:

1. «Дискретная математика и теоретическая информатика»

и обеспечивает изучение последующих дисциплин:

1. «Сети ЭВМ»

3.3 Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения образовательной программы обучающийся должен достичь следующие результаты обучения по дисциплине:

Код компетенции/ индикатора компетенции	Наименование компетенции/индикатора компетенции
ПК-1	Способен осуществлять проведение научно-исследовательских и опытно-конструкторских разработок по отдельным разделам темы
ПК-1.2	<i>Осуществляет выполнение экспериментов и оформление результатов исследований и разработок</i>
ПК-3	Способен разрабатывать требования и проектировать программное обеспечение
ПК-3.1	<i>Анализирует требования к программному обеспечению</i>

4 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1 Содержание разделов дисциплины

4.1.1 Наименование тем и часы на все виды нагрузки

№ п/п	Наименование темы дисциплины	Лек, ач	Пр, ач	ИКР, ач	СР, ач
1	Введение	2	4	1	3
2	Группировка и цензурирование исходных данных	2	3		3
3	Разведочный анализ	4	3		3
4	Регрессионный анализ	6	3	0	3
5	Корреляционный анализ	3	3		3
6	Дисперсионный анализ	3	3		3
7	Многомерное нормальное распределение	3	3		3
8	Методы снижения размерности	4	3		3
9	Регрессия на главные компоненты	3	3		3
10	Кластерный анализ	2	3		6
11	Дискриминантный анализ	2	3		6
	Итого, ач	34	34	1	39
	Из них ач на контроль	0	0	0	0
	Общая трудоемкость освоения, ач/зе			108/3	

4.1.2 Содержание

№ п/п	Наименование темы дисциплины	Содержание
1	Введение	Предмет, цель, задачи и основные проблемы анализа данных. Инструменты с открытым кодом, обеспечивающие анализ данных. Концепция практикума, поддерживающего курс по изучению и освоению инструментов для анализа данных
2	Группировка и цензурирование исходных данных	Одномерная группировка. Одномерное цензурирование. Таблицы сопряженности. Гипотеза независимости. Гипотеза однородности. Поле корреляции. Многомерные группировки. Многомерное цензурирование. Робастность в многомерном статистическом анализе. Проверка наблюдений на аномальность.

№ п/п	Наименование темы дисциплины	Содержание
3	Разведочный анализ	Идеология разведочного анализа, выявление точек-выбросов, проверка однородности дисперсий ,проверка нормальности распределения данных, выявление избыточного количества нулевых значений, выявление коллинеарных переменных, выявление характера связи между анализируемыми переменными, выявление взаимодействий между переменными-предикторами, выявление пространственно-временных корреляций между значениями зависимой переменной;
4	Регрессионный анализ	Основные задачи регрессионного анализа. Выбор адекватного уравнения регрессии. Парная регрессия. Множественная регрессия. Линейная множественная регрессионная модель. Понятие о нелинейной регрессии. Теорема Маркова. Оценка значимости уравнения регрессии и остаточной дисперсии с помощью метода наименьших квадратов. Корреляционный анализ многомерной генеральной совокупности, его назначение и место. Логистическая регрессия
5	Корреляционный анализ	Корреляционный анализ количественных связей и порядковых переменных, катеризованные корреляции. Оценка корреляционной матрицы. Оценки частных и множественных коэффициентов корреляции. Проверка существенности связи. Методы изучения связи. Непараметрические показатели связи. Непараметрический подход, основанный на частотах. Непараметрический подход, основанный на рангах. Преимущества и недостатки непараметрических методов. Проверка взаимосвязи между качественными признаками. Показатели тесноты связи. Коэффициент Фехнера (знаков). Коэффициент Спирмена. Коэффициент конкордации. Коэффициенты ассоциации и контингенции. Коэффициент взаимной сопряженности Пирсона.
6	Дисперсионный анализ	Сущность дисперсионного анализа. Условия применения дисперсионного анализа. Однофакторный дисперсионный анализ. Внутригрупповая и межгрупповая вариация. Гипотезы для дисперсионного анализа. Двухфакторный дисперсионный анализ. Многомерный дисперсионный анализ.
7	Многомерное нормальное распределение	Многомерное нормальное распределение, как основная модель современных многомерных методов. Случайный вектор. Свойства многомерного нормального распределения.. Устойчивость многомерного нормального распределение относительно линейных преобразований. Доверительные и толерантные множества. Проблема размерности в многомерных .исследованиях. Многомерные методы оценивания и статистического сравнения.

№ п/п	Наименование темы дисциплины	Содержание
8	Методы снижения размерности	Модель, математическое обоснование и алгоритм метода главных компонент. Собственные векторы и собственные значения корреляционной матрицы и их использование для получения матрицы весовых коэффициентов. Главные компоненты. Главные компоненты двумерного, трехмерного и конечномерного пространства. Линейная модель факторного анализа. Различие предпосылок компонентного и факторного анализа. Основные проблемы факторного анализа. Факторное отображение и факторная структура. Компоненты дисперсии в факторном анализе. Преобразование корреляционной матрицы в факторном анализе. Получение и использование матрицы нагрузок и матрицы индивидуальных значений. Метод главных факторов. Получение первого главного фактора. Использование методов вращения. Формирование названия общего фактора.
9	Регрессия на главные компоненты	Получение и использование матрицы индивидуальных значений главных компонент. Особенности регрессии на главные компоненты. Регрессия на общие факторы. Сравнение результатов компонентного и факторного анализа. Использование идей факторного анализа в компонентном.
10	Кластерный анализ	Задача многомерной классификации объектов исследования. Классификация без обучения. Кластерный анализ. Методы кластерного анализа. Расстояние между объектами. Меры близости между объектами. Меры близости между кластерами. Классификация признаков на основе матриц коэффициентов статистической связи между ними. Иерархические кластер-процедуры. Метод К-средних. Классификация больших совокупностей объектов методами параллельных процедур. Классификация в пространстве главных компонент и общих факторов. Функционалы качества разбиения на классы. Зависимость выбора метода классификации от цели исследования.
11	Дискриминантный анализ	Математическое описание метода дискриминантного анализа. Обучающие выборки. Линейный дискриминантный анализ при известных параметрах многомерного нормального закона распределения. Вероятность ошибочной классификации с помощью дискриминантной функции. Оценка качества дискриминантной функции и информативности отдельных признаков. Пошаговый дискриминантный анализ.

4.2 Перечень лабораторных работ

Лабораторные работы не предусмотрены.

4.3 Перечень практических занятий

Наименование практических занятий	Количество ауд. часов
1. Установка программ с открытым кодом для выполнения анализа данных	4
2. Группировка и цензурирование исходных данных	3
3. Разведочный анализ	3
4. Регрессионный анализ	3
5. Корреляционный анализ	3
6. Дисперсионный анализ	3
7. Многомерное нормальное распределение	3
8. Методы снижения размерности	3
9. Регрессия на главные компоненты	3
10. Кластерный анализ	3
11. Дискриминантный анализ	3
Итого	34

4.4 Курсовое проектирование

Курсовая работа (проект) не предусмотрены.

4.5 Реферат

Реферат не предусмотрен.

4.6 Индивидуальное домашнее задание

Индивидуальное домашнее задание не предусмотрено.

4.7 Доклад

Доклад не предусмотрен.

4.8 Кейс

Кейс не предусмотрен.

4.9 Организация и учебно-методическое обеспечение самостоятельной работы

Изучение дисциплины сопровождается самостоятельной работой студентов с рекомендованными преподавателем литературными источниками и информационными ресурсами сети Интернет.

Планирование времени для изучения дисциплины осуществляется на весь период обучения, предусматривая при этом регулярное повторение пройденного материала. Обучающимся, в рамках внеаудиторной самостоятельной работы, необходимо регулярно дополнять сведениями из литературных источников материал, законспектированный на лекциях. При этом на основе изучения рекомендованной литературы целесообразно составить конспект основных положений, терминов и определений, необходимых для освоения разделов учебной дисциплины.

Особое место уделяется консультированию, как одной из форм обучения и контроля самостоятельной работы. Консультирование предполагает особым образом организованное взаимодействие между преподавателем и студентами, при этом предполагается, что консультант либо знает готовое решение, которое он может предписать консультируемому, либо он владеет способами деятельности, которые указывают путь решения проблемы.

Текущая СРС	Примерная трудоемкость, ач
Работа с лекционным материалом, с учебной литературой	14
Опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	0
Самостоятельное изучение разделов дисциплины	0
Выполнение домашних заданий, домашних контрольных работ	0
Подготовка к лабораторным работам, к практическим и семинарским занятиям	14
Подготовка к контрольным работам, коллоквиумам	0
Выполнение расчетно-графических работ	0
Выполнение курсового проекта или курсовой работы	0

Текущая СРС	Примерная трудоемкость, ач
Поиск, изучение и презентация информации по заданной проблеме, анализ научных публикаций по заданной теме	0
Работа над междисциплинарным проектом	0
Анализ данных по заданной теме, выполнение расчетов, составление схем и моделей, на основе собранных данных	0
Подготовка к зачету, дифференцированному зачету, экзамену	11
ИТОГО СРС	39

5 Учебно-методическое обеспечение дисциплины

5.1 Перечень основной и дополнительной литературы, необходимой для освоения дисциплины

№ п/п	Название, библиографическое описание	К-во экз. в библ.
Основная литература		
1	Есипов, Борис Алексеевич. Методы исследования операций : учеб. пособие для вузов по направлению ВПО 010300 "Фундам. информатика и информац. технологии" / Б. А. Есипов, 2013. -299 с.	19
Дополнительная литература		
1	Таха, Хэмди А. Введение в исследование операций : монография / Х.А.Таха; [Пер.с англ. В.Ю.Тюпти, А.А.Минько], 2001. -911 с.	10

5.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», используемых при освоении дисциплины

№ п/п	Электронный адрес
1	The Super Decisions https://superdecisions.com/

5.3 Адрес сайта курса

Адрес сайта курса: <https://vec.etu.ru/moodle/course/view.php?id=17888>

6 Критерии оценивания и оценочные материалы

6.1 Критерии оценивания

Для дисциплины «Анализ данных в программах с открытым кодом» предусмотрены следующие формы промежуточной аттестации: зачет.

Зачет

На трёх контрольных точках студенты предоставляют отчёты по Контрольным Работам (КР).

Материалы КР связаны с обработкой информации, по методам, изложенными в лекционном материале.

Результаты КР оцениваются по критериям, приведенным ниже:

0 баллов - работа выполнена с большим количеством существенных ошибок, требующих большого количества времени на исправление. Студент испытывает серьезные трудности при ответе на вопросы по теоретической части работы.

3 балла - работа содержит несколько существенных ошибок, требующих большого количества времени на исправление. Студент испытывает серьезные трудности при ответе на некоторые из вопросов по теоретической части работы.

4 балла - работа содержит ряд незначительных ошибок, которые могут быть исправлены в короткие сроки. Студент испытывает затруднения при ответе на некоторые из вопросов по теоретической части работы.

5 баллов - работа содержит одну или две незначительные ошибки, которые могут быть исправлены в короткие сроки, или выполнена верно. Студент не испытывает затруднений при ответе на вопросы по теоретической части работы.

В конце семестра подсчитывается суммарное число баллов, набранных по результатам выполнения трёх КР.

При наборе 12-ти и более суммарных баллов студент получает «Зачёт» по дисциплине.

Если число суммарных баллов меньше 12-ти баллов -студент получает «Незачёт» по дисциплине и обязан в сроки, устанавливаемые деканатом, выполнить корректировку КР и пройти новую процедуру оценивания.

Особенности допуска

Допуск к зачету: выполнение всех практических работ на практических занятиях.

6.2 Оценочные материалы для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине

Образцы задач (заданий) для контрольных (проверочных) работ

Примеры заданий контрольных работ

Задание 1. Привести разделение на кластеры на $n = 10$ объектов, каждый из которых характеризуется признаками x, y, z , со значениями:

$$x = 252, 322, 339, 406, 492, 597, 695, 751, 903, 971,$$

$$y = 2100, 2700, 2000, 1000, 2200, 2300, 2300, 2100, 2400, 2400,$$

$$z = 2600, 2600, 2000, 1000, 2000, 2000, 2400, 2100, 2100, 2500.$$

Задание 2. Представлены данные медицинского обследования 12-ти пациентов:

$$\text{по Весу в кг} = 97, 73, 85, 100, 87, 70, 71, 105, 97, 69, 65, 82,$$

$$\text{по Росту в см} = 184, 170, 173, 187, 172, 165, 155, 178, 175, 160, 151, 181,$$

$\text{по Показателю Давления} = 120, 130, 125, 127, 140, 131, 128, 135, 120, 110, 115, 125.$

Построить столбцовую диаграмму для переменной Вес.

Построить диаграмму рассеяния для переменных Вес и Рост.

Задание 3. Проверить необходимыми расчётами можно ли считать, что средние двух нормальных совокупностей равны, если выборочные средние и дисперсии, вычисленные по двум выборкам объёмом $n_1 = 16, n_2 = 9$ равны $ср1 = 12,57, ср2 = 11,87$, $дисп1 = 0,91, дисп2 = 0,95$. Предполагается, что дисперсии

обеих совокупностей равны. Уровень значимости принять равным 0,05.

Задание 4. Составить таблицу квантилей порядка 0,01 0,05 0,1 0,9 0,95 0,99 для распределения $N(0,1)$, Стюдента с $k=10$ и Хи квадрат с $k=19$.

Задание 5. Привести по три варианта признаков, измеренных в Номинальной, Ранговой, Интервальной и Абсолютной шкалах. Перечислите дескриптивные статистики, допустимые в каждой из шкал.

Весь комплект контрольно-измерительных материалов для проверки сформированности компетенции (индикатора компетенции) размещен в закрытой части по адресу, указанному в п. 5.3

6.3 График текущего контроля успеваемости

Неделя	Темы занятий	Вид контроля
1	Группировка и цензурирование исходных данных Разведочный анализ Регрессионный анализ	Контрольная работа
2		
3		
4		
5	Корреляционный анализ Дисперсионный анализ Многомерное нормальное распределение Методы снижения размерности	Контрольная работа
6		
7		
8		
9	Регрессия на главные компоненты Кластерный анализ Дискриминантный анализ	Контрольная работа
10		
11		
12		
13		Контрольная работа

6.4 Методика текущего контроля

на лекционных занятиях

Текущий контроль включает в себя контроль посещаемости (не менее **80** % занятий), по результатам которого студент получает допуск на зачет.

на практических (семинарских) занятиях

Текущий контроль включает в себя контроль посещаемости (не менее **80** % занятий), по результатам которого студент получает допуск на зачет.

В ходе проведения семинарских и практических занятий целесообразно привлечение студентов к как можно более активному участию в дискуссиях, решении задач, обсуждениях и т. д. При этом активность студентов также может учитываться преподавателем, как один из способов текущего контроля на практических занятиях.

Результаты контрольных работ оцениваются по критериям, приведенным ниже:

0 баллов - работа выполнена с большим количеством существенных ошибок

бок, требующих большого количества времени на исправление. Студент испытывает серьезные трудности при ответе на вопросы по теоретической части работы.

3 балла - работа содержит несколько существенных ошибок, требующих большого количества времени на исправление. Студент испытывает серьезные трудности при ответе на некоторые из вопросов по теоретической части работы.

4 балла - работа содержит ряд незначительных ошибок, которые могут быть исправлены в короткие сроки. Студент испытывает затруднения при ответе на некоторые из вопросов по теоретической части работы.

5 баллов - работа содержит одну или две незначительные ошибки, которые могут быть исправлены в короткие сроки, или выполнена верно. Студент не испытывает затруднений при ответе на вопросы по теоретической части работы.

В конце семестра подсчитывается суммарное число баллов, набранных по результатам выполнения трёх КР.

самостоятельной работы студентов

Контроль самостоятельной работы студентов осуществляется на лекционных и практических занятиях студентов по методикам, описанным выше.

7 Описание информационных технологий и материально-технической базы

Тип занятий	Тип помещения	Требования к помещению	Требования к программному обеспечению
Лекция	Лекционная аудитория	Количество посадочных мест – в соответствии с контингентом, рабочее место преподавателя, доска, экран, проектор, ноутбук	1) Windows XP и выше; 2) Microsoft Office 2007 и выше
Практические занятия	Аудитория	Количество посадочных мест – в соответствии с контингентом, рабочее место преподавателя, доска, экран, проектор, ноутбук	1) Windows XP и выше; 2) Microsoft Office 2007 и выше
Самостоятельная работа	Помещение для самостоятельной работы	Оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.	1) Windows XP и выше; 2) Microsoft Office 2007 и выше

8 Адаптация рабочей программы для лиц с ОВЗ

Адаптированная программа разрабатывается при наличии заявления со стороны обучающегося (родителей, законных представителей) и медицинских показаний (рекомендациями психолого-медико-педагогической комиссии). Для инвалидов адаптированная образовательная программа разрабатывается в соответствии с индивидуальной программой реабилитации.

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

№ п/п	Дата	Изменение	Дата и номер протокола заседания УМК	Автор	Начальник ОМОЛА