

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Галунин Сергей Александрович
Должность: проректор по учебной работе
Дата подписания: 27.07.2022 12:30:40
Уникальный программный ключ:
08ef34338325bdb0ac5a47baa5472ce36cc3fc3b

Приложение к ОПОП
«Автономные интеллектуальные
системы»



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 И ИНТЕРПРЕТАЦИЯ ДАННЫХ»

для подготовки магистров

по направлению

09.04.04 «Программная инженерия»

по программе

«Автономные интеллектуальные системы»

Санкт-Петербург

2021

ЛИСТ СОГЛАСОВАНИЯ

Разработчики:

д.т.н., профессор Геппенер В.В.

Рабочая программа рассмотрена и одобрена на заседании кафедры МОЭВМ
07.09.2021, протокол № 7

Рабочая программа рассмотрена и одобрена учебно-методической комиссией
ФКТИ, 16.09.2021, протокол № 6

Согласовано в ИС ИОТ

Начальник ОМОЛА Загороднюк О.В.

1 СТРУКТУРА ДИСЦИПЛИНЫ

Обеспечивающий факультет	ФКТИ
Обеспечивающая кафедра	МОЭВМ
Общая трудоемкость (ЗЕТ)	4
Курс	2
Семестр	3
Виды занятий	
Лекции (академ. часов)	17
Практические занятия (академ. часов)	34
Иная контактная работа (академ. часов)	1
Все контактные часы (академ. часов)	52
Самостоятельная работа, включая часы на контроль (академ. часов)	92
Всего (академ. часов)	144
Вид промежуточной аттестации	
Дифф. зачет (курс)	2

2 АННОТАЦИЯ ДИСЦИПЛИНЫ

«АНАЛИЗ И ИНТЕРПРЕТАЦИЯ ДАННЫХ»

В курсе рассматриваются вопросы алгоритмического обеспечения систем, предназначенных для анализа и интерпретации данных. Описываются методы анализа данных и дальнейшей интерпретации полученных результатов. Существенное внимание уделено вопросам классификации данных с использованием детерминированных и статистических моделей. Рассмотрены методы снижения размерностей данных. Изучаются новые методы анализа данных на основе технологии Data Mining. Анализируются современные пакеты прикладных программ для решения задач обработки экспериментальных данных.

SUBJECT SUMMARY

«DATA ANALYSIS AND INTERPRETATION»

The course deals with the algorithmic software systems for data analysis and interpretation. It describes the methods of data analysis and further interpretation of the results. Considerable attention is paid to issues of data classification using deterministic and statistical models. The methods of reducing data dimensions are considered. We study new methods of data analysis based on Data Mining technology. Analyzed modern software packages for the solution of data processing tasks.

3 ОБЩИЕ ПОЛОЖЕНИЯ

3.1 Цели и задачи дисциплины

1. Слушатель должен свободно ориентироваться в вопросах алгоритмического обеспечения систем, предназначенных для анализа и интерпретации данных.
2. Овладение методами анализа данных и дальнейшей интерпретацией полученных результатов. Умение разбираться в вопросах классификации данных с использованием детерминированных и статистических моделей, методах снижения размерностей данных, новых методах анализа данных на основе технологии Data Mining и современных пакетах прикладных программ для решения задач обработки экспериментальных данных.
3. Изучение основных методов решения на ЭВМ задач анализа и интерпретации данных, получаемых с помощью различного рода информационно-измерительных систем. Знание базовых алгоритмов анализа и интерпретации данных.
4. Формирование навыков применения методов решения на ЭВМ задач анализа и интерпретации данных при разработке алгоритмов анализа и обработки измерительной информации. Умение использовать стандартную терминологию, определения и обозначения
5. Освоение основных тенденций развития теории и практики анализа и интерпретации данных. Приобретение практических навыков работы с современными пакетами прикладных программ для решения задач анализа и интерпретации данных.

3.2 Место дисциплины в структуре ОПОП

Дисциплина изучается на основе ранее освоенных дисциплин учебного плана:

1. «Анализ, моделирование и оптимизация систем»

2. «Математические методы распознавания образов»

3. «Нейронные сети»

4. «Машинное обучение»

и обеспечивает подготовку выпускной квалификационной работы.

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-11	Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач
<i>ПК-11.2</i>	<i>Разрабатывает унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизма контроля за соблюдением указанных методологий</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	1			
2	Тема 1. Основные понятия дисциплины	1			10
3	Тема 2. Классификация данных с использованием детерминированных моделей	2	6		12
4	Тема 3. Классификация данных на основе статистических моделей	2	6		12
5	Тема 4. Кластер-анализ	2	6		12
6	Тема 5. Методы снижения размерностей данных	2	6		12
7	Тема 6. Системы DATA MINING в задачах анализа и интерпретации данных	4	6		12
8	Тема 7. Современные пакеты прикладных программ для решения задач обработки экспериментальных данных	2	4		10
9	Заключение	1		1	12
	Итого, ач	17	34	1	92
	Из них ач на контроль	0	0	0	0
	Общая трудоемкость освоения, ач/зе	144/4			

4.1.2 Содержание

№ п/п	Наименование темы дисциплины	Содержание
1	Введение	Предмет дисциплины и ее задачи. Краткие сведения о развитии теории анализа и интерпретации данных. Структура, содержание дисциплины, ее связи с другими дисциплинами учебного плана специальности. Перечень дисциплин и их разделов, усвоение которых необходимо студентам для изучения курса «Анализ и интерпретация данных».
2	Тема 1. Основные понятия дисциплины	Введение в анализ данных. Проблема обработки данных. Матрица данных. Гипотезы компактности и скрытых факторов. Структура матрицы данных и задачи обработки. Матрица объект-объект и признак-признак. Расстояние и близость. Измерение признаков. Отношения и их представление. Основные проблемы измерений. Основные типы шкал. Проблема адекватности. Основные задачи анализа и интерпретации данных.

№ п/п	Наименование темы дисциплины	Содержание
3	Тема 2. Классификация данных с использованием детерминированных моделей	Решающие поверхности и дискриминантные функции. Линейные дискриминантные функции. Линейная разделимость. Кусочно-линейные дискриминантные функции. Нелинейные дискриминантные функции. Фи-машины. Потенциальные функции как дискриминантные функции. Пространство весов. Процедуры обучения с коррекцией ошибок: правило с фиксированным приращением, правило абсолютной коррекции, частично корректирующее правило. Обобщенные градиентные методы. Персептронный критерий. Процедуры обучения на основе минимальной среднеквадратичной ошибки: псевдоинверсный метод, метод Хо-Кашпа
4	Тема 3. Классификация данных на основе статистических моделей	Функция потерь. Байесовская дискриминантная функция. Принятие решение по максимуму правдоподобия. Оптимальная дискриминантная функция для нормально распределенных образов. Дискриминантная функция Фишера. Множественный дискриминантный анализ. Пошаговый дискриминантный анализ. Ошибки классификации. Примеры построения статистических дискриминантных функций для различных статистических моделей данных. Обучение для статистических дискриминантных функций. Оценки максимального правдоподобия, байесовские оценки. Непараметрическое оценивание. Парзеновские окна, метод непараметрического оценивания на основе К-ближайшего соседства
5	Тема 4. Кластер-анализ	Основные типы задач кластер-анализа. Меры подобия и функции расстояния. Выбор критерия кластеризации. Кластерные методы, основанные на евклидовой метрике. Иерархическая кластеризация. Метод К-внутригрупповых средних. Использование методов теории графов в задачах кластеризации. Кластеризация на основе анализа плотностей вероятностей

№ п/п	Наименование темы дисциплины	Содержание
6	Тема 5. Методы снижения размерностей данных	Анализ матриц исходных данных. Метод главных компонент. Корреляционная матрица и ее основные свойства. Собственные векторы и собственные числа корреляционной матрицы. Приведение корреляционной матрицы к диагональной форме. Геометрическая интерпретация главных компонент на плоскости. Модели факторного анализа. Оценка факторных нагрузок методом максимального правдоподобия и центроидным методом. Вращение факторов и их интерпретация. Использование кластеризации признаков для снижения размерности. Многомерное шкалирование. Метрический и неметрический подход к многомерному шкалированию. Методы ортогонального проектирования. Нелинейные методы многомерного шкалирования. Многомерное шкалирование неметрических данных. Многомерные развертки
7	Тема 6. Системы DATA MINING в задачах анализа и интерпретации данных	Понятие об интеллектуальных системах анализа и интерпретации данных. DATA MINING -системы извлечения новых знаний из данных. Типы систем DATA MINING -предметно-ориентированные аналитические системы, статистические пакеты, нейронные сети, деревья решений, обнаружение логических закономерностей, генетические алгоритмы, системы визуализации многомерных данных
8	Тема 7. Современные пакеты прикладных программ для решения задач обработки экспериментальных данных	Табличные процессоры и базы данных в задачах обработки данных. Виды статистических пакетов. Требования к статистическим пакетам общего назначения. Общая характеристика пакетов "STATGRAFICS Plus", "STATISTICA", SAS, SPSS . Комплексные системы класса DATA MINING для обработки данных - "PolyAnalyst", Intelligent Miner
9	Заключение	Перспективные направления развития и использования методов анализа и интерпретации данных

4.2 Перечень лабораторных работ

Лабораторные работы не предусмотрены.

4.3 Перечень практических занятий

Наименование практических занятий	Количество ауд. часов
1. Изучение приемов работы с системой STATISTICA	4
2. Основы работы с языком STATISTICA Basic	4
3. Знакомство с пакетом программ STATISTICA и подготовка статистических данных	4

Наименование практических занятий	Количество ауд. часов
4. Изучение дискриминантного анализа с помощью пакета программ STATISTICA	4
5. Исследование методов кластер-анализа	4
6. Работа в среде пакета обучения и распознавания ГЕКОНАЛ	4
7. Исследование методов факторного анализа	5
8. Исследование методов многомерного шкалирования	5
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	20
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	20
Подготовка к контрольным работам, коллоквиумам	22
Выполнение расчетно-графических работ	0
Выполнение курсового проекта или курсовой работы	0
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	30
ИТОГО СРС	92

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 библи.
Основная литература		
1	Технологии анализа данных: Data Mining, Text Mining, Visual Mining, OLAP. 2 изд. [Электронный ресурс] / А. Барсегян, М. Куприянов, В. Степаненко и др., 2008. -384 с.	неогр
2	Барсегян А. А. Методы и модели анализа данных: OLAP и Data Mining [Электронный ресурс] / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод, 2004. -336 с.	неогр
3	Вуколов, Эдуард Александрович. Основы статистического анализа [Текст] : практикум по статистическим методам и исследованию операций с использованием пакетов STATISTIKA и EXSEL : учеб. пособие по специальности "Менеджмент организации" / Э.А. Вуколов, 2004. -462 с.	17
Дополнительная литература		
1	Дюк В. Data Mining [Текст] : Учеб. курс / В.Дюк, А.Самойленко, 2001. - 366 с.	21
2	Дэйвисон, Марк Л. Многомерное шкалирование: Методы наглядного представления данных [Текст] / М.Л. Дэйвисон; Пер. с англ. В.С.Каменского, 1988. -254 с. с.	4
3	Айвазян, Сергей Артемьевич. Прикладная статистика: Основы моделирования и первичная обработка данных [Текст] : Справ. изд. / С.А.Айвазян, И.С.Енюков , Л.Д.Мешалкин, 1983. -471 с.	15
4	Айвазян, Сергей Артемьевич. Прикладная статистика: Исследование зависимостей [Текст] : Справ. изд. / С.А.Айвазян, И.С.Енюков, Л.Д.Мешалкин; Под ред. С.А.Айвазяна, 1985. -487 с.	13

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

№ п/п	Электронный адрес
1	Курс Data Mining на Интуит - http://www.intuit.ru/studies/courses/6/6/info

5.3 Адрес сайта курса

Адрес сайта курса: <https://vec.etu.ru/moodle/course/view.php?id=7348>

6 Критерии оценивания и оценочные материалы

6.1 Критерии оценивания

Для дисциплины «Анализ и интерпретация данных» формой промежуточной аттестации является зачет с оценкой.

Зачет с оценкой

Оценка	Описание
Неудовлетворительно	Курс не освоен. Студент испытывает серьезные трудности при ответе на ключевые вопросы дисциплины
Удовлетворительно	Студент в целом овладел курсом, но некоторые разделы освоены на уровне определений и формулировок теорем
Хорошо	Студент овладел курсом, но в отдельных вопросах испытывает затруднения. Умеет решать задачи
Отлично	Студент демонстрирует полное овладение курсом, способен применять полученные знания при решении конкретных задач.

Особенности допуска

Для допуска к дифференцированному зачету необходимо получить не менее "Удовлетворительно" по каждой контрольной точке в рамках текущего контроля успеваемости.

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

Примерные вопросы к дифф.зачету

№ п/п	Описание
1	Проблема обработки данных. Матрица данных
2	Нелинейные дискриминантные функции. Фи-машины
3	Функция потерь. Байесовская дискриминантная функция.
4	Оценки максимального правдоподобия, байесовские оценки
5	Иерархическая кластеризация. Метод K-внутригрупповых средних
6	Модели факторного анализа. Оценка факторных нагрузок методом максимального правдоподобия и центроидным методом

Форма билета

Дисциплина **Анализ и интерпретация данных**

1. Оценки максимального правдоподобия, байесовские оценки
2. Модели факторного анализа. Оценка факторных нагрузок методом максимального правдоподобия и центроидным методом

Образцы задач (заданий) для контрольных (проверочных) работ

Контрольные работы состоят в развернутом письменном ответе на вопросы. На каждой контрольной работе студент должен решить задачу.

Пример задачи: построить байесовское решающее правило для двух классов для нормального распределения с заданными параметрами.

Пример контрольных вопросов:

1. Линейные дискриминантные функции.
2. Байесовская дискриминантная функция.
4. Примеры построения статистических дискриминантных функций для различных статистических моделей данных.
5. Метод главных компонент.

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	Тема 1. Основные понятия дисциплины	
2	Тема 2. Классификация данных с использованием детерминированных моделей Тема 3. Классификация данных на основе статистических моделей Тема 4. Кластер-анализ	
3		
4		
5		
6		
7		
8		Контрольная работа
9	Тема 5. Методы снижения размерностей данных Тема 6. Системы DATA MINING в задачах анализа и интерпретации данных	
10		
11		
12		
13		
14		
15		
16		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 дифференцированный зачет.

на практических (семинарских) занятиях

Текущий контроль включает в себя контроль посещаемости (не менее **80** % занятий), а также 2 контрольные работы, по результатам которых студент получает допуск на дифференцированный зачет.

Контрольные работы должны быть выполнены не менее, чем на оценку ”удовлетворительно”.

Оценивание контрольных работ проводится по трем критериям:

- 1) для решения задачи использованы правильные формулы;

2) в ходе решения отсутствуют ошибки;

3) дано пошаговое описание хода решения.

При выполнении всех трех критериев ставится оценка "Отлично", при выполнении критериев 1 и 2 ставится оценка "Хорошо", при выполнении только критерия 1 ставится оценка "Удовлетворительно".

В ходе проведения 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.

самостоятельной работы студентов

Контроль самостоятельной работы студентов осуществляется на лекционных, лабораторных 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, персональный компьютер IBM-совместимый Pentium или выше, проектор, экран, меловая или маркерная доска	1) Windows 7 и выше; 2) Microsoft Office 2007 и выше;
Практические занятия	Аудитория	Количество посадочных мест, оборудованных компьютерами IBM-совместимыми Pentium или выше, – в соответствии с контингентом, рабочее место преподавателя, проектор, экран, меловая или маркерная доска, персональный компьютер IBM совместимый Pentium или выше	1) Windows 7 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