

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Галунин Сергей Александрович
Должность: проректор по учебной работе
Дата подписания: 21.06.2023 09:58:01
Уникальный программный ключ:
08ef34338325bdb0ac5a47baa5472ce36cc3fc3b

Приложение к ОПОП
«Управление и информатика в
технических системах»



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
«РАСПРЕДЕЛЕННАЯ ОБРАБОТКА ДАННЫХ»
для подготовки бакалавров
по направлению
27.03.04 «Управление в технических системах»
по профилю
«Управление и информатика в технических системах»

Санкт-Петербург

2022

ЛИСТ СОГЛАСОВАНИЯ

Разработчики:

к.т.н., доцент каф. АПУ Каплун Д.И.

Рабочая программа рассмотрена и одобрена на заседании кафедры АПУ
18.05.2022, протокол № 1

Рабочая программа рассмотрена и одобрена учебно-методической комиссией
ФКТИ, 24.02.2022, протокол № 2

Согласовано в ИС ИОТ

Начальник ОМОЛА Загороднюк О.В.

1 СТРУКТУРА ДИСЦИПЛИНЫ

Обеспечивающий факультет	ФКТИ
Обеспечивающая кафедра	АПУ
Общая трудоемкость (ЗЕТ)	2
Курс	4
Семестр	8

Виды занятий

Лекции (академ. часов)	6
Практические занятия (академ. часов)	8
Иная контактная работа (академ. часов)	2
Все контактные часы (академ. часов)	16
Самостоятельная работа, включая часы на контроль (академ. часов)	56
Всего (академ. часов)	72

Вид промежуточной аттестации

Дифф. зачет (семестр) 8

2 АННОТАЦИЯ ДИСЦИПЛИНЫ

«РАСПРЕДЕЛЕННАЯ ОБРАБОТКА ДАННЫХ»

Освещаемые в курсе теоретические и прикладные вопросы полезны для студентов, специализирующихся в области управления и информационных технологий в технических системах, компьютерного и математического моделирования. В курсе даются основы анализа данных, включающие вопросы выявления аномалий в сырых данных, классификации и кластеризации, корреляционного и регрессионного анализа, машинного обучения и искусственного интеллекта. Кроме того, в курсе рассматриваются такие современные программно-аппаратные технологии распределённой обработки данных, как облачные и туманные вычисления, грид-вычисления, вычисления на основе GPU и FPGA. Уделяется внимание решению практических задач с использованием современных программно-аппаратных средств (MATLAB, Python, GPU, FPGA).

SUBJECT SUMMARY

«DISTRIBUTED DATA PROCESSING»

The theoretical and applied questions covered in the course are useful for students specializing in applied mathematics and computer science, computer and mathematical modeling. The course provides the basics of data analysis, including anomaly detection in raw data, classification and clustering, correlation and regression analysis, machine learning and artificial intelligence. In addition, the course discusses such modern software and hardware technologies for distributed data processing as cloud and fog computing, grid computing, GPU and FPGA based computing. Attention is paid to solving practical problems using modern software and hardware (MATLAB, Python, GPU, FPGA).

3 ОБЩИЕ ПОЛОЖЕНИЯ

3.1 Цели и задачи дисциплины

1. Цели освоения дисциплины «Распределенная обработка данных» состоят в формировании у студентов теоретических знаний в области предназначения и использования распределенных систем для обработки информации и практических навыков построения распределенных систем различными программными средствами.

2. Задачи дисциплины:

1). Знания основ анализа данных, включающих вопросы выявления аномалий в сырьих данных, классификации и кластеризации, корреляционного и регрессионного анализа, машинного обучения и искусственного интеллекта.

2). Формирование профессиональных навыков обработки данных с использованием методов и алгоритмов выявления аномалий в сырьих данных, классификации и кластеризации, корреляционного и регрессионного анализа, машинного обучения и искусственного интеллекта.

3). Освоение современных программно-аппаратных средств распределённой обработки и визуализации данных.

3. После изучения дисциплины студент должен обладать знаниями по:

-методам распределенной обработки данных, синхронизацию процессов через доступ к общим ресурсам, понятие о критических интервалах, семафорах, программирование параллельных

алгоритмов с помощью критических интервалов и семафоров;

-методам распределенного параллельного программирования, синхронизацию процессов, взаимодействие распределенных процессов;

-методам искусственного интеллекта.

4. Должен обладать умениями:

-создавать параллельные программы для алгоритмов матричных задач, для задач решения систем линейных уравнений, для задач сортировок разными методами, для алгоритмов матричных задач на вычислительных системах с топологией двух и трех мерных решеток;

-использовать методы искусственного интеллекта.

5. Должен владеть навыками:

-разработки параллельных программ для алгоритмов матричных задач, для задач решения систем линейных уравнений, для задач сортировок разными методами, для алгоритмов матричных задач на вычислительных системах с топологией двух и трех мерных решеток;

-интеллектуальной обработки данных.

3.2 Место дисциплины в структуре ОПОП

Дисциплина изучается на основе ранее освоенных дисциплин учебного плана:

1. «Дискретная математика и теоретическая информатика»
2. «Интернет-технологии»

и обеспечивает изучение последующих дисциплин:

1. «Информационные технологии в управлении»

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-0	Способен разрабатывать информационные модели и применять их для решения задач профессиональной деятельности
ПК-0.1	<i>Знает современные виды информационных моделей, применяемых при решении задач профессиональной деятельности</i>
ПК-7	Способен проводить вычислительные эксперименты с использованием стандартных программных средств с целью получения математических моделей процессов и объектов автоматизации и управления
ПК-7.2	<i>Умеет разрабатывать математические модели процессов и объектов управления на основе экспериментальных данных</i>
ПК-8	Готов участвовать в подготовке и внедрению результатов разработок в производство средств и систем управления
ПК-8.4	<i>Участвует в подготовке и внедрении результатов разработок в производство средств и систем управления</i>
ПК-11	Способен владеть методами проектирования распределенных систем управления с применением современных информационных технологий
ПК-11.2	<i>Умеет применять информационные технологии для решения задач синтеза и анализа распределенных систем управления</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	0.5	0.5	1	5
2	Программно-аппаратные технологии и инструменты распределённой обработки данных	0.5	0.5	0	6
3	Выявление аномалий в сырых данных	0.5	0.5	0	6
4	Классификация	0.5	0.5	0	6
5	Кластеризация	0.5	1	0	6
6	Частотный анализ	0.5	1	0	6
7	Определение временных задержек и корреляционный анализ	0.5	1	0	5
8	Регрессионный анализ	0.5	1	0	5
9	Машинное обучение и искусственный интеллект	1	1	0	6
10	Заключение	1	1	1	5
	Итого, ач	6	8	2	56
	Из них ач на контроль	0	0	0	4
	Общая трудоемкость освоения, ач/зе			72/2	

4.1.2 Содержание

№ п/п	Наименование темы дисциплины	Содержание
1	Введение	Наука о данных (Data Science). Распределённая обработка данных. Исторический экскурс. Содержание области знаний.
2	Программно-аппаратные технологии и инструменты распределённой обработки данных	Параллельные и распределённые алгоритмы обработки данных. Распределённые операционные системы. Многопоточность. Многозадачность. Архитектуры параллельных и распределённых вычислительных систем. Облачные вычисления. Туманные вычисления. Грид-вычисления. Вычисления на основе GPU и FPGA. Векторные вычисления.

№ п/п	Наименование темы дисциплины	Содержание
3	Выявление аномалий в сырых данных	Сырые данные. Временные ряды. Выбросы. Причины выбросов. Определение выбросов. Гистограмма. Ящик с усами. Критерии выбросов. Критерий Шовене. Критерий Граббса. Критерий Пирса. Критерий Диксона. Правило 3 сигм. Статистические тесты. Модельные тесты. Итерационные методы. Метрические методы. Методы подмены задачи. Методы машинного обучения. Ансамбли алгоритмов. Оценка качества определения выбросов.
4	Классификация	Метод ближайших соседей. Байесовский классификатор. Деревья решений. Метод главных компонент. Машина опорных векторов. Нейронная сеть. Бустинг. Бэггинг.
5	Кластеризация	Алгоритм k-средних. Алгоритм k-медиан. Метод нечёткой кластеризации С-средних. Нейронные сети Кохонена. Генетический алгоритм. Иерархическая кластеризация.
6	Частотный анализ	Сырые данные. Временные ряды. Компоненты временных рядов. Объединение компонентов временного ряда. Оценка тренда. Виды функциональных зависимостей. Выделение тренда. МНК. Метод скользящего среднего. Метод экспоненциального сглаживания. Прогнозирование. Двойное экспоненциальное сглаживание. Метод Хольт-Винтерса. Качество прогнозирования.
7	Определение временных задержек и корреляционный анализ	Системы с запаздыванием. Метод ближайших соседей. Корреляция и корреляционный анализ. Коэффициенты корреляции. Коэффициент Пирсона. Коэффициент Спирмена. Коэффициент Кендалла. Коэффициент Фехнера. RS-анализ. Показатель Херста. Ограничения корреляционного анализа.
8	Регрессионный анализ	Регрессия и регрессионный анализ. Метод наименьших квадратов. Линейная и полиномиальная модель. Обобщенный МНК.
9	Машинное обучение и искусственный интеллект	Нейронная сеть (НС). Что такое НС? Классификация НС. Области применения НС. Структура нейрона. Синапс. Принцип работы НС. Функция активации. Тренировочный сет. Эпоха. Ошибка. Нейрон смещения. Обучение НС. Градиентный спуск. Метод обратного распространения. Гиперпараметры. Сходимость. Переобучение. Свёрточные нейронные сети (СНС). Архитектура СНС. Свёрточные слои. Субдискретизирующие слои. Пуллинг. Полносвязная НС. Карты.
10	Заключение	Современное состояние области знаний. Перспективы.

4.2 Перечень лабораторных работ

Лабораторные работы не предусмотрены.

4.3 Перечень практических занятий

Наименование практических занятий	Количество ауд. часов
1. Выявление аномалий в сырых данных.	1
2. Классификация.	1
3. Кластеризация.	1
4. Частотный анализ.	1
5. Определение временных задержек.	1
6. Корреляционный анализ.	1
7. Регрессионный анализ.	1
8. Машинное обучение и искусственный интеллект.	1
Итого	8

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

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Текущая СРС	Примерная трудоемкость, ач
Работа с лекционным материалом, с учебной литературой	18
Опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	8
Самостоятельное изучение разделов дисциплины	10
Выполнение домашних заданий, домашних контрольных работ	0
Подготовка к лабораторным работам, к практическим и семинарским занятиям	10
Подготовка к контрольным работам, коллоквиумам	6
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, дифференциированному зачету, экзамену	4
ИТОГО СРС	56

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	Мартин, Джеймс. Вычислительные сети и распределенная обработка данных. Программное обеспечение, методы и архитектура : [в 2 вып.] [Текст]. Вып. 2, 1986. -269 с.	90
2	Обработка данных в среде MATLAB [Текст] : метод. указания к практическим занятиям / Санкт-Петербургский государственный электротехнический университет им. В.И. Ульянова (Ленина) "ЛЭТИ", 2013. -33, [1] с.	34
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	Тироф Р. Обработка данных в управлении [Текст]. Т. 2 / В.Я.Алтаев, 1976. -534 с. с.	9
5	Обработка данных в среде Excel [Текст] : метод. указания к лабораторным работам / Санкт-Петербургский государственный электротехнический университет им. В.И. Ульянова (Ленина) "ЛЭТИ", 2011. -27, [1] с.	13
Дополнительная литература		
1	Берикашвили, Валерий Шалвович. Статистическая обработка данных, планирование эксперимента и случайные процессы [Электронный ресурс] : Учебное пособие для вузов / Берикашвили В. Ш., Оськин С. П., 2021. -164 с	неогр.
2	Пселтис Э. Д. Потоковая обработка данных. Конвейер реального времени [Электронный ресурс], 2018. -218 с.	неогр.

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

№ п/п	Электронный адрес
1	Факультет аналитики Big Data https://geekbrains.ru/geek_university/big-data-analytics
2	Data Science https://vk.com/datasience
3	Школа анализа данных https://yandexdataschool.ru/
4	MathWorks MATLAB & Simulink https://vk.com/mathworks
5	MATHWORKS https://matlab.ru/
6	Select a Web Site https://www.mathworks.com/

№ п/п	Электронный адрес
7	MATLAB: инструмент будущего или дорогая игрушка https://geekbrains.ru/posts/how_to_matlab
8	MachineLearning.ru http://www.machinelearning.ru
9	Coursera https://www.coursera.org

5.3 Адрес сайта курса

Адрес сайта курса: <https://vec.etu.ru/moodle/course/view.php?id=7570>

6 Критерии оценивания и оценочные материалы

6.1 Критерии оценивания

Для дисциплины «Распределенная обработка данных» предусмотрены следующие формы промежуточной аттестации: 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
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 дифф. зачету необходимо выполнить и защитить все практические работы, выполнить контрольную работу на оценку "удовлетворительно" и выше. Обязательно посещение не менее 80% лекций. Оценка за дифференцированный зачет выставляется по результатам текущей аттестации в ходе семестра.

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

Вопросы к дифф.зачету

№ п/п	Описание
1	История и предпосылки создания Java
2	Особенности языка Java. Лексические основы (зарезервированные ключевые слова, идентификаторы, литеры, разделители).
3	Типы данных и операции языка Java.
4	Автоматическое преобразование типов в выражениях.
5	Управляющие операторы
6	Понятие и описание массивов. Элементы и длина массива.
7	Принципы объектно-ориентированного программирования. Достоинства и недостатки ООП
8	Синтаксис описания класса, метода. Объявление объекта. Передача параметров, перегрузка методов
9	Конструкторы класса. Перегрузка конструкторов. Вложенные, внутренние, абстрактные классы
10	Технологии распределенной обработки данных
11	Характеристика технологии клиент-сервер
12	Модель клиент-сервер
13	Связь посредством сообщений. Сохранность и синхронность во взаимодействиях. Нерезидентная связь на основе сообщений. Сохранная связь на основе сообщений
14	Связь на основе потоков данных. Поддержка непрерывных сред. Потоки данных и качество обслуживания. Синхронизация потоков данных
15	Клиенты. Пользовательские интерфейсы. Клиентское программное обеспечение, обеспечивающее прозрачность распределения.
16	Серверы. Общие вопросы разработки. Серверы объектов.
17	Работа с базами данных. Модели доступа к данным. Описание ODBC, JDBC
18	Классификация подходов к построению распределенной информационной системы.
19	Основные задачи, выполняемые серверными программами

20	Технологии построения распределенных информационных систем: технологии CORBA и COM
----	--

Образцы задач (заданий) для контрольных (проверочных) работ

Пример контрольной работы

1. Перечислите способы параллельной обработки данных.
2. Опишите организацию интеграции приложений на уровне прикладных процессов.
3. Опишите структура интернет/интранет портала системы обработки данных.
4. Перечислите способы последовательной обработки данных.
5. Опишите организацию интеграции приложений на уровне пользователя.

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	Классификация	
2		
3		
4		
5		Практическая работа
6	Кластеризация	
7		
8		
9		
10		
11	Машинное обучение и искусственный интеллект	Практическая работа
12		
13		
14		
15		
16	Контрольная работа	
17		

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 дифф. зачет.

Текущий контроль материала в рамках лекционной подготовки осуществляется также в виде контрольной работы. Для этого разработан тест, охватывающий некоторые темы дисциплины.

Максимальное число баллов, которые можно набрать за контрольную работу - 10.

При оценке каждого вопроса используется следующая шкала:

-2 балл – ответ верный;

-1 балл – ответ частично верный;

- 0 баллов – ответ неверный, или ответ отсутствует.

Критерии оценки контрольной работы:

8 – 10 баллов – «отлично»;

6 – 7 баллов – «хорошо»;

4– 5 баллов – «удовлетворительно»;

менее 4 баллов – «неудовлетворительно».

на практических занятиях

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

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 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, проектор, экран, ноутбук	1) Windows XP и выше; 2) Microsoft Office 2007 и выше
Практические занятия	Аудитория	Количество посадочных мест и ПК – в соответствии с контингентом, рабочее место преподавателя, доска.	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