

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение высшего
образования
"САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ"

Кафедра № 82

УТВЕРЖДАЮ
Руководитель направления
проф. д.пед.н., доц.
(должность, уч. степень, звание)

А.Г. Степанов
(инициалы, фамилия)
(подпись)
24 июня 2021 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Методы обработки больших данных»
(Наименование дисциплины)

Код направления подготовки/ специальности	09.03.03
Наименование направления подготовки/ специальности	Прикладная информатика
Наименование направленности	Прикладная информатика в экономике
Форма обучения	очная

Лист согласования рабочей программы дисциплины

Программу составил (а)
доцент, к.т.н.
(должность, уч. степень, звание) (подпись, дата) (инициалы, фамилия)

Программа одобрена на заседании кафедры № 82
19 мая 2021 г, протокол № 10

Заведующий кафедрой № 82
д.э.н., доц.
(уч. степень, звание) (подпись, дата) (инициалы, фамилия)

Ответственный за ОП ВО 09.03.03(03)
Ст. преподаватель
(должность, уч. степень, звание) (подпись, дата) (инициалы, фамилия)

Заместитель директора института №8 по методической работе
доц. к.э.н., доц.
(должность, уч. степень, звание) (подпись, дата) (инициалы, фамилия)

Аннотация

Дисциплина «Методы обработки больших данных» входит в образовательную программу высшего образования по направлению подготовки/ специальности 09.03.03 «Прикладная информатика» направленности «Прикладная информатика в экономике». Дисциплина реализуется кафедрой «№82».

Дисциплина нацелена на формирование у выпускника следующих компетенций:

ПК-4 «способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач»

ПК-5 «Способность принимать участие в организации ИТ-инфраструктуры и управлении информационной безопасностью»

Содержание дисциплины охватывает круг вопросов, связанных с методами решения задач обработки и анализа больших данных, принципы построения современных систем хранения данных, возможности высокопроизводительных вычислительных систем, технологии распределенных вычислений, методы и модели Data Mining..

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, лабораторные работы.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточная аттестация в форме зачета.

Общая трудоемкость освоения дисциплины составляет 3 зачетных единицы, 108 часов.

Язык обучения по дисциплине «русский»

1. Перечень планируемых результатов обучения по дисциплине

1.1. Цели преподавания дисциплины

Формирование у студентов необходимой теоретической базы и практических навыков, которые позволят всесторонне и системно понимать современные проблемы прикладной математики и информатики, проблемы обработки и анализа информации, а также разрабатывать и анализировать концептуальные и теоретические модели при решении научных и прикладных задач в области информационных технологий. Одна из главных проблем современной обработки и анализа данных - рост объемов данных, поэтому вопросам обработки большого объема данных посвящена данная дисциплина. Главная задача курса - сформировать целостное представление о современных проблемах анализа и обработки больших данных, помочь овладеть опытом разработки и анализа концептуальных и теоретических моделей прикладных задач анализа больших данных с применением моделей Data Mining.

1.2. Дисциплина входит в состав части, формируемой участниками образовательных отношений, образовательной программы высшего образования (далее – ОП ВО).

1.3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения ОП ВО.

В результате изучения дисциплины обучающийся должен обладать следующими компетенциями или их частями. Компетенции и индикаторы их достижения приведены в таблице 1.

Таблица 1 – Перечень компетенций и индикаторов их достижения

Категория (группа) компетенции	Код и наименование компетенции	Код и наименование индикатора достижения компетенции
Профессиональные компетенции	ПК-4 способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач	ПК-4.В.1 владеть навыками ведения БД и решения прикладных задачи с использованием современных информационных технологий поддержки информационного обеспечения при решении прикладных задач
Профессиональные компетенции	ПК-5 Способность принимать участие в организации ИТ-инфраструктуры и управлении информационной безопасностью	ПК-5.3.1 знать варианты проектных решений по видам информационных систем

2. Место дисциплины в структуре ОП

Дисциплина может базироваться на знаниях, ранее приобретенных обучающимися при изучении следующих дисциплин:

- « Основы интеллектуального анализа».
- «Технологии программирования».
- « Операционные системы».
- «Интеллектуальные информационные системы»,

Знания, полученные при изучении материала данной дисциплины, имеют как самостоятельное значение, так и могут использоваться при изучении других дисциплин:

- «Интеллектуальный анализ данных»

3. Объем и трудоемкость дисциплины

Данные об общем объеме дисциплины, трудоемкости отдельных видов учебной работы по дисциплине (и распределение этой трудоемкости по семестрам) представлены в таблице 2.

Таблица 2 – Объем и трудоемкость дисциплины

Вид учебной работы	Всего	Трудоемкость по семестрам
		№8
1	2	3
Общая трудоемкость дисциплины, ЗЕ/ (час)	3/ 108	3/ 108
Аудиторные занятия, всего час.	30	30
в том числе:		
лекции (Л), (час)	10	10
практические/семинарские занятия (ПЗ), (час)		
лабораторные работы (ЛР), (час)	20	20
курсовой проект (работа) (КП, КР), (час)		
экзамен, (час)		
Самостоятельная работа, всего (час)	78	78
Вид промежуточной аттестации: зачет, дифф. зачет, экзамен (Зачет, Дифф. зач, Экз.**)	Зачет	Зачет

Примечание: ** кандидатский экзамен

4. Содержание дисциплины

4.1. Распределение трудоемкости дисциплины по разделам и видам занятий.

Разделы, темы дисциплины и их трудоемкость приведены в таблице 3.

Таблица 3 – Разделы, темы дисциплины, их трудоемкость

Разделы, темы дисциплины	Лекции (час)	ПЗ (СЗ) (час)	ЛР (час)	КП (час)	СРС (час)
Семестр 8					
Раздел 1. Технологии анализа данных Тема 1.1. Большие данные (Big Data): современные подходы к обработке и хранению. Проблема множественного сравнения данных.	1		2		8
Тема 1.2. Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных.	1		2		8

Раздел 2. Методы многомерного статистического анализа и анализа нечисловой информации Тема 2.1. Начала корреляционного анализа и регрессионного анализа	2		4		14
Раздел 3. Программные средства для обработки данных и системы Data Mining Тема 3.1. Программные инструменты интеллектуального анализа данных на примере пакета RapidMiner	2		8		32
Раздел 4. Аналитика в больших данных Тема 4.1. Обзор задач и методов машинного обучения и интеллектуального анализа больших данных	2		2		8
Тема 4.2. Интерактивное исследование данных	2		2		8
Итого в семестре:	10		20		78
Итого	10	0	20	0	78

4.2. Содержание разделов и тем лекционных занятий.

Содержание разделов и тем лекционных занятий приведено в таблице 4.

Таблица 4 – Содержание разделов и тем лекционного цикла

Номер раздела	Название и содержание разделов и тем лекционных занятий
1	Раздел 1. Технологии анализа данных Тема 1.1. Большие данные (Big Data): современные подходы к обработке и хранению. Проблема множественного сравнения данных. Тема 1.2. Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных.
2	Раздел 2. Методы многомерного статистического анализа и анализа нечисловой информации Тема 2.1. Начала корреляционного анализа и регрессионного анализа
3	Раздел 3. Программные средства для обработки данных и системы Data Mining Тема 3.1. Программные инструменты интеллектуального анализа данных на примере пакета RapidMiner
4	Раздел 4. Аналитика в больших данных Тема 4.1. Обзор задач и методов машинного обучения и интеллектуального анализа больших данных Тема 4.2. Интерактивное исследование данных

4.3. Практические (семинарские) занятия

Темы практических занятий и их трудоемкость приведены в таблице 5.

Таблица 5 – Практические занятия и их трудоемкость

№ п/п	Темы практических занятий	Формы практических занятий	Трудоемкость, (час)	№ раздела дисциплины
Учебным планом не предусмотрено				
Всего				

4.4. Лабораторные занятия

Темы лабораторных занятий и их трудоемкость приведены в таблице 6.

Таблица 6 – Лабораторные занятия и их трудоемкость

№ п/п	Наименование лабораторных работ	Трудоемкость, (час)	№ раздела дисциплины
Семестр 8			
1	Начальное знакомство со средой обработки больших данных. Методы подготовки исходных данных	4	1
2	Исследование операций преобразования данных.	4	2
3	Метод деревьев принятия решений для анализа больших данных	4	4
4	Методы регрессии для анализа больших данных.	4	3
5	Методы нейронных сетей для анализа больших данных	4	3
Всего		20	

4.5. Курсовое проектирование/ выполнение курсовой работы

Учебным планом не предусмотрено

4.6. Самостоятельная работа обучающихся

Виды самостоятельной работы и ее трудоемкость приведены в таблице 7.

Таблица 7 – Виды самостоятельной работы и ее трудоемкость

Вид самостоятельной работы	Всего, час	Семестр 8, час
1	2	3
Изучение теоретического материала дисциплины (ТО)	40	40
Курсовое проектирование (КП, КР)		
Расчетно-графические задания (РГЗ)		
Выполнение реферата (Р)	20	20
Подготовка к текущему контролю успеваемости (ТКУ)	4	4
Домашнее задание (ДЗ)	10	10
Контрольные работы заочников (КРЗ)		
Подготовка к промежуточной аттестации (ПА)	4	4
Всего:	78	78

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
Учебно-методические материалы для самостоятельной работы обучающихся указаны в п.п. 7-11.

6. Перечень печатных и электронных учебных изданий
Перечень печатных и электронных учебных изданий приведен в таблице 8.
Таблица 8– Перечень печатных и электронных учебных изданий

Шифр/ URL адрес	Библиографическая ссылка	Количество экземпляров в библиотеке (кроме электронных экземпляров)
004.5 Б 71	Блюм В.С. Методы интеллектуального анализа данных (пакет программ Rapid Miner) : учебное пособие / В. С. Блюм ; С.-Петербург. гос. ун-т аэрокосм. приборостроения. - СПб. : Изд-во ГУАП, 2018. - 205 с. - ISBN 978-5-8088-1304-5 :	20
004.6 Б 77	Андреас Вайгенд. BIG DATA. ВСЯ ТЕХНОЛОГИЯ В ОДНОЙ КНИГЕ, 2017, 416 с	5
005.2 Б 67	Бн Анналин, Су Кеннет, Теоретический минимум по Big Data. Всё, что нужно знать о больших данных. — СПб.: Питер, 2019. — 208 с	5

7. Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»
Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины приведен в таблице 9.
Таблица 9 – Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

URL адрес	Наименование
https://geekbrains.ru/geek_university/big-data-analytics?utm_source=yandex.ru&utm_medium=pc&utm_campaign=yandex_s_spb_big_data_main_lp_big_data_prog&utm_term=big%20data&utm_content=cid:41888689 gid:3746106369 aid:7223943732 pt:premium pos:1 st:search src:none dvc:desktop reg:2 adph:no apt:none ret: dsa:16210584538&yclid=3435126429319716288	Факультет аналитики Big Data
https://rb.ru/howto/chto-takoe-big-data/	Что такое Big data: собрали всё самое важное о больших данных
https://mining-cryptocurrency.ru/big-data/	Что такое Big Data простыми словами? Применение и перспективы больших данных

8. Перечень информационных технологий

8.1. Перечень программного обеспечения, используемого при осуществлении образовательного процесса по дисциплине.

Перечень используемого программного обеспечения представлен в таблице 10.

Таблица 10– Перечень программного обеспечения

№ п/п	Наименование
1	Программный пакет RapidMiner
2	Microsoft Office Excel
3	Microsoft Office Word
4	Microsoft Office Power Point

8.2. Перечень информационно-справочных систем, используемых при осуществлении образовательного процесса по дисциплине

Перечень используемых информационно-справочных систем представлен в таблице 11.

Таблица 11– Перечень информационно-справочных систем

№ п/п	Наименование
1	http://www. Yandex.ru

9. Материально-техническая база

Состав материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине, представлен в таблице 12.

Таблица 12 – Состав материально-технической базы

№ п/п	Наименование составной части материально-технической базы	Номер аудитории (при необходимости)
1	Дисплейный класс	14-06 – 14-11 Ленс.
2	Мультимедийная лекционная аудитория	

10. Оценочные средства для проведения промежуточной аттестации

10.1. Состав оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине приведен в таблице 13.

Таблица 13 – Состав оценочных средств для проведения промежуточной аттестации

Вид промежуточной аттестации	Перечень оценочных средств
Зачет	Вопросы по темам/разделам дисциплины
Задания для самостоятельной работы	Комплект заданий

10.2. В качестве критериев оценки уровня сформированности (освоения) компетенций обучающимися применяется 5-балльная шкала оценки сформированности компетенций, которая приведена в таблице 14. В течение семестра может использоваться 100-балльная шкала модульно-рейтинговой системы Университета, правила использования которой, установлены соответствующим локальным нормативным актом ГУАП.

Таблица 14 –Критерии оценки уровня сформированности компетенций

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	

Оценка компетенции 5-балльная шкала	Характеристика сформированных компетенций
«отлично» «зачтено»	<ul style="list-style-type: none"> – обучающийся глубоко и всесторонне усвоил программный материал; – уверенно, логично, последовательно и грамотно его излагает; – опираясь на знания основной и дополнительной литературы, тесно привязывает усвоенные научные положения с практической деятельностью направления; – умело обосновывает и аргументирует выдвигаемые им идеи; – делает выводы и обобщения; – свободно владеет системой специализированных понятий.
«хорошо» «зачтено»	<ul style="list-style-type: none"> – обучающийся твердо усвоил программный материал, грамотно и по существу излагает его, опираясь на знания основной литературы; – не допускает существенных неточностей; – увязывает усвоенные знания с практической деятельностью направления; – аргументирует научные положения; – делает выводы и обобщения; – владеет системой специализированных понятий.
«удовлетворительно» «зачтено»	<ul style="list-style-type: none"> – обучающийся усвоил только основной программный материал, по существу излагает его, опираясь на знания только основной литературы; – допускает несущественные ошибки и неточности; – испытывает затруднения в практическом применении знаний направления; – слабо аргументирует научные положения; – затрудняется в формулировании выводов и обобщений; – частично владеет системой специализированных понятий.
«неудовлетворительно» «не зачтено»	<ul style="list-style-type: none"> – обучающийся не усвоил значительной части программного материала; – допускает существенные ошибки и неточности при рассмотрении проблем в конкретном направлении; – испытывает трудности в практическом применении знаний; – не может аргументировать научные положения; – не формулирует выводов и обобщений.

10.3. Типовые контрольные задания или иные материалы.

Вопросы (задачи) для экзамена представлены в таблице 15.

Таблица 15 – Вопросы (задачи) для экзамена

№ п/п	Перечень вопросов (задач) для экзамена
	Учебным планом не предусмотрено

Вопросы (задачи) для зачета / дифф. зачета представлены в таблице 16.

Таблица 16 – Вопросы (задачи) для зачета / дифф. зачета

№ п/п	Перечень вопросов (задач) для зачета / дифф. зачета
1	История возникновения проблемы «больших данных»
2	Причины возникновения проблемы «больших данных» и условия её решения
3	Характеристики больших данных
4	Сравнительные характеристики обычных и больших данных
5	Методика A/B testing.
6	Методика Ad-hoc GRID
7	Методики Association rule learning
8	Методика BOINC-грид
9	Технология Data GRID
10	Методики Data fusion and data integration

11	Когнитивная кластеризация
12	Метод Ensemble learning.
13	Методика Genetic algorithms
14	Инфраструктура ГРИД
15	Вычислительная система со множественным потоком команд и множественным потоком данных
16	Методики распознавания естественного языка человека.
17	Методики анализа связей между узлами в сетях.
18	Численные методы для редизайна сложных систем и процессов
19	Набор методик с элементами самообучения для предсказания поведенческой модели потребителей.
20	Набор статистических методов для выявления закономерности
21	Методики оценки настроений потребителей на основе технологии распознавания естественного языка человека
22	Методики распознавания сигнала на фоне шума и его дальнейшего анализа.
23	Методики на основе технологий машинного обучения, позволяющие выявить функциональные взаимосвязи в анализируемых массивах данных.
24	Методика темпоральной логики анализа повторяющихся с течением времени последовательностей данных.
25	Методы графического представления результатов интеллектуального анализа данных в виде диаграмм или анимированных изображений

Перечень тем для курсового проектирования/выполнения курсовой работы представлены в таблице 17.

Таблица 17 – Перечень тем для курсового проектирования/выполнения курсовой работы

№ п/п	Примерный перечень тем для курсового проектирования/выполнения курсовой работы
	Учебным планом не предусмотрено

Вопросы для проведения промежуточной аттестации в виде тестирования представлены в таблице 18.

Таблица 18 – Примерный перечень вопросов для тестов

№ п/п	Примерный перечень вопросов для тестов
1	Учебным планом не предусмотрено

Перечень тем контрольных работ по дисциплине обучающихся заочной формы обучения, представлены в таблице 19.

Таблица 19 – Перечень контрольных работ

№ п/п	Перечень контрольных работ
	Не предусмотрено

10.4. Методические материалы, определяющие процедуры оценивания индикаторов, характеризующих этапы формирования компетенций, содержатся в локальных нормативных актах ГУАП, регламентирующих порядок и процедуру проведения текущего контроля успеваемости и промежуточной аттестации обучающихся ГУАП.

11. Методические указания для обучающихся по освоению дисциплины

11.1. Методические указания для обучающихся по освоению лекционного материала .

Основное назначение лекционного материала – логически стройное, системное, глубокое и ясное изложение учебного материала. Назначение современной лекции в рамках дисциплины не в том, чтобы получить всю информацию по теме, а в освоении

фундаментальных проблем дисциплины, методов научного познания, новейших достижений научной мысли. В учебном процессе лекция выполняет методологическую, организационную и информационную функции. Лекция раскрывает понятийный аппарат конкретной области знания, её проблемы, дает цельное представление о дисциплине, показывает взаимосвязь с другими дисциплинами.

Планируемые результаты при освоении обучающимися лекционного материала:

- получение современных, целостных, взаимосвязанных знаний, уровень которых определяется целевой установкой к каждой конкретной теме;
- получение опыта творческой работы совместно с преподавателем;
- развитие профессионально-деловых качеств, любви к предмету и самостоятельного творческого мышления.
- появление необходимого интереса, необходимого для самостоятельной работы;
- получение знаний о современном уровне развития науки и техники и о прогнозе их развития на ближайшие годы;
- научиться методически обрабатывать материал (выделять главные мысли и положения, приходить к конкретным выводам, повторять их в различных формулировках);
- получение точного понимания всех необходимых терминов и понятий.

Лекционный материал может сопровождаться демонстрацией слайдов и использованием раздаточного материала при проведении коротких дискуссий об особенностях применения отдельных тематик по дисциплине.

Структура предоставления лекционного материала:

- Лекционный материал представлен в форме презентаций пяти лекций, размещенных в курсе «Методы обработки больших данных» в системе дистанционного обучения ЛМС,
- Дополнительный материал к каждой лекции, представленный в форме файлов Word, размещенных в курсе «Методы обработки больших данных» в системе дистанционного обучения ЛМС.

11.2. Методические указания для обучающихся по участию в семинарах *(не предусмотрено)*

11.3. Методические указания для обучающихся по прохождению практических занятий. Не предусмотрено

11.4. Методические указания для обучающихся по выполнению лабораторных работ

В ходе выполнения лабораторных работ обучающийся должен углубить и закрепить знания, практические навыки, овладеть современной методикой и техникой эксперимента в соответствии с квалификационной характеристикой обучающегося. Выполнение лабораторных работ состоит из экспериментально-практической, расчетно-аналитической частей и контрольных мероприятий.

Выполнение лабораторных работ обучающимся является неотъемлемой частью изучения дисциплины, определяемой учебным планом, и относится к средствам, обеспечивающим решение следующих основных задач обучающегося:

- приобретение навыков исследования процессов, явлений и объектов, изучаемых в рамках данной дисциплины;
- закрепление, развитие и детализация теоретических знаний, полученных на лекциях;
- получение новой информации по изучаемой дисциплине;
- приобретение навыков самостоятельной работы с лабораторным оборудованием и приборами.

Задание и требования к проведению лабораторных работ
Представлены на странице курса в <http://lms.guap.ru>.

Структура и форма отчета о лабораторной работе
Представлены на странице курса в <http://lms.guap.ru>.

Требования к оформлению отчета о лабораторной работе
Представлены на странице курса в <http://lms.guap.ru>.

11.5. Методические указания для обучающихся по прохождению курсового проектирования/выполнения курсовой работы

Не предусмотрено

11.6. Методические указания для обучающихся по прохождению самостоятельной работы

В ходе выполнения самостоятельной работы, обучающийся выполняет работу по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

Для обучающихся по заочной форме обучения, самостоятельная работа может включать в себя контрольную работу.

В процессе выполнения самостоятельной работы, у обучающегося формируется целесообразное планирование рабочего времени, которое позволяет им развивать умения и навыки в усвоении и систематизации приобретаемых знаний, обеспечивает высокий уровень успеваемости в период обучения, помогает получить навыки повышения профессионального уровня.

Методическими материалами, направляющими самостоятельную работу обучающихся являются:

– учебно-методический материал по дисциплине;

Представлены на странице курса в <http://lms.guap.ru>.

11.7. Методические указания для обучающихся по прохождению текущего контроля успеваемости.

Текущий контроль успеваемости предусматривает контроль качества знаний обучающихся, осуществляемого в течение семестра с целью оценивания хода освоения дисциплины.

Для проведения текущего контроля успеваемости на каждом лекционном занятии выделяется до 10 минут, а также контроль успеваемости проводится при приеме реферата и лабораторных работ. Результаты текущего контроля успеваемости учитываются при проведении промежуточной аттестации.

11.8. Методические указания для обучающихся по прохождению промежуточной аттестации.

Промежуточная аттестация обучающихся предусматривает оценивание промежуточных и окончательных результатов обучения по дисциплине. Она включает в себя:

– зачет – это форма оценки знаний, полученных обучающимся в ходе изучения учебной дисциплины в целом или промежуточная (по окончании семестра) оценка знаний обучающимся по отдельным разделам дисциплины с аттестационной оценкой «зачтено» или «не зачтено».

Лист внесения изменений в рабочую программу дисциплины

Дата внесения изменений и дополнений. Подпись внесшего изменения	Содержание изменений и дополнений	Дата и № протокола заседания кафедры	Подпись зав. кафедрой