

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Воронежский государственный технический университет»

УТВЕРЖДАЮ
Декан факультета информационных
технологий и компьютерной безопасности
П.Ю. Гусев
«31» августа 2021 г.

РАБОЧАЯ ПРОГРАММА
дисциплины

«Методы обработки больших данных»

Направление подготовки 09.03.01 Информатика и вычислительная техника

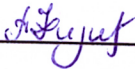
Профиль Системы автоматизированного проектирования

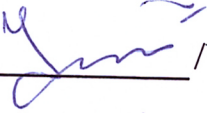
Квалификация выпускника бакалавр


Нормативный период обучения 4 года / 4 года и 11 м.

Форма обучения очная / заочная

Год начала подготовки 2021

Автор программы  /Филимонова А.А./

Заведующий кафедрой
Компьютерных
интеллектуальных
технологий проектирования  /Чижов М.И./

Руководитель ОПОП  /Бредихин А.В./

Воронеж 2021

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

1.1. Цели дисциплины

Цель учебной дисциплины заключается в получении обучающимися теоретических знаний о процессах управления с последующим применением в профессиональной сфере и развитии практических навыков в области науки об управлении, с последующим применением в профессиональной сфере, формирование способностей решать средствами математики задачи управления.

Углубить изучение методологических и теоретических проблем, связанных с выявлением устойчивых, повторяющихся связей в социально-экономических процессах, их структурных характеристик, закономерностей функционирования и тенденций развития экономических отношений, объяснением на этой основе существующих факторов и феноменов.

1.2. Задачи освоения дисциплины

1. Усвоение знаний о сущности, структуре и видах математических моделей принятия решений;

2. Формирование представлений о содержании, формах, особенностях дисциплины «Технология работы с большими данными»;

3. Развитие навыков создания и решения моделей, необходимых в сфере управления.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Методы обработки больших данных» относится к дисциплинам части, формируемой участниками образовательных отношений (дисциплина по выбору) блока Б1.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Процесс изучения дисциплины «Методы обработки больших данных» направлен на формирование следующих компетенций:

ПК-5 - Способен обеспечивать производственный процесс машиностроительного предприятия программным обеспечением в соответствии с предъявляемыми требованиями

ПК-3 - Способен применять методы моделирования в профессиональной деятельности

Компетенция	Результаты обучения, характеризующие сформированность компетенции
ПК-5	знать методы и средства системного анализа, оптимизации, управления, принятия решений и обработки информации в машиностроении
	уметь разрабатывать методы проектирования и анализа алгоритмов, используемых в производственных процессах
	владеть современными методами исследования и информационно-коммуникационными технологиями

ПК-3	знать математические методы обработки данных
	уметь выполнять исследования процессов создания, накопления и обработки информации, включая анализ и создание моделей данных и знаний, языков их описания и манипулирования
	владеть новыми методами исследования и обработки данных и их применению в самостоятельной научно-исследовательской деятельности в области профессиональной деятельности

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины «Методы обработки больших данных» составляет 5 з.е.

Распределение трудоемкости дисциплины по видам занятий
очная форма обучения

Виды учебной работы	Всего часов	Семестры
		6
Аудиторные занятия (всего)	108	108
В том числе:		
Лекции	36	36
Лабораторные работы (ЛР)	72	72
Самостоятельная работа	45	45
Курсовая работа	+	+
Часы на контроль	27	27
Виды промежуточной аттестации - экзамен	+	+
Общая трудоемкость:		
академические часы	180	180
зач.ед.	5	5

заочная форма обучения

Виды учебной работы	Всего часов	Семестры
		8
Аудиторные занятия (всего)	24	24
В том числе:		
Лекции	8	8
Лабораторные работы (ЛР)	16	16
Самостоятельная работа	147	147
Курсовая работа	+	+
Часы на контроль	9	9
Виды промежуточной аттестации - экзамен	+	+
Общая трудоемкость:		
академические часы	180	180
зач.ед.	5	5

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1 Содержание разделов дисциплины и распределение трудоемкости по видам занятий

очная форма обучения

№ п/п	Наименование темы	Содержание раздела	Лекц	Лаб. зан.	СРС	Всего, час
1	Введение, основные понятия анализа данных	Терминология, история появления. Технические сложности работы с большими данными. Распределенная файловая система HDFS.	6	12	6	24
2	Модель вычислений MapReduce	Модель вычислений MapReduce. Реализация MapReduce в Hadoop.	6	12	8	26
3	Автоматизированная загрузка и обработка больших данных	Hadoop как технология хранения и обработки больших данных. Способы загрузки данных в Hadoop. Базовые операции с данными.	6	12	8	26
4	Обзор задач и методов машинного обучения и интеллектуального анализа больших данных	Аналитические методы, применимые к большим объемам данных. Многомерные связи, ассоциации, корреляции. Непрерывность поступающих обновлений как характерная черта анализа больших данных. Примеры обработки неструктурированных данных.	6	12	8	26
5	Оценивание качества алгоритмов и моделей	Метрики оценивания ошибки обучения алгоритмов. Мониторинг качества моделей. Создание шаблонов построения моделей. Автоматическое построение моделей по сегментам данных.	6	12	8	26
6	Особенности реальных больших данных	Исследование реальных больших данных. Примеры обработки реальных данных.	6	12	7	25
Итого			36	72	45	153

заочная форма обучения

№ п/п	Наименование темы	Содержание раздела	Лекц	Лаб. зан.	СРС	Всего, час
1	Введение, основные понятия анализа данных	Терминология, история появления. Технические сложности работы с большими данными. Распределенная файловая система HDFS.	2	4	24	30
2	Модель вычислений MapReduce	Модель вычислений MapReduce. Реализация MapReduce в Hadoop.	2	4	24	30
3	Автоматизированная загрузка и обработка больших данных	Hadoop как технология хранения и обработки больших данных. Способы загрузки данных в Hadoop. Базовые операции с данными.	2	2	24	28
4	Обзор задач и методов машинного обучения и интеллектуального анализа больших данных	Аналитические методы, применимые к большим объемам данных. Многомерные связи, ассоциации, корреляции. Непрерывность поступающих обновлений как характерная черта анализа больших данных. Примеры обработки неструктурированных данных.	2	2	24	28
5	Оценивание качества алгоритмов и моделей	Метрики оценивания ошибки обучения алгоритмов. Мониторинг качества моделей. Создание шаблонов построения моделей. Автоматическое построение моделей по сегментам данных.	-	2	26	28
6	Особенности реальных больших данных	Исследование реальных больших данных. Примеры обработки реальных данных.	-	2	25	27
Итого			8	16	147	171

5.2 Перечень лабораторных работ

1. Анализ больших данных, представленных в табличной форме.
2. Проведение анализа больших данных по открытому набору данных.
3. Визуальный углубленный анализ больших данных.
4. Разработка модели, использующей линейную регрессию, для классификации входного потока больших данных.
5. Подготовка отчета по анализу набора больших данных.

6. ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ ПРОЕКТОВ (РАБОТ) И КОНТРОЛЬНЫХ РАБОТ

В соответствии с учебным планом освоение дисциплины предусматривает выполнение курсовой работы в 6 семестре для очной формы обучения, в 8 семестре для заочной формы обучения.

Примерная тематика курсовой работы: «Обработка входного потока больших данных»

Задачи, решаемые при выполнении курсовой работы:

- Сбор и анализ исходных данных
- Визуальный анализ исходных данных
- Разработка модели-классификатора
- Подготовка отчета по проделанной работе

Курсовая работа включает в себя графическую часть и расчетно-пояснительную записку.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

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

7.1.1 Этап текущего контроля

Результаты текущего контроля знаний и межсессионной аттестации оцениваются по следующей системе:

«аттестован»;

«не аттестован».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Аттестован	Не аттестован
ПК-5	знать методы и средства системного анализа, оптимизации, управления, принятия решений и обработки информации в машиностроении	Контрольная работа	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	уметь разрабатывать методы проектирования и анализа алгоритмов, используемых в производственных процессах	Контрольная работа	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	владеть современными методами исследования и информационно-коммуникационными технологиями	Контрольная работа	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
ПК-3	знать математические методы обработки данных	Контрольная работа	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	уметь выполнять исследования процессов создания, накопления и обработки информации, включая анализ и создание моделей данных и знаний, языков их описания и манипулирования	Контрольная работа	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах

владеть новыми методами исследования и обработки данных и их применению в самостоятельной научно-исследовательской деятельности в области профессиональной деятельности	Контрольная работа	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
---	--------------------	---	---

7.1.2 Этап промежуточного контроля знаний

Результаты промежуточного контроля знаний оцениваются в 6 семестре для очной формы обучения, 8 семестре для заочной формы обучения по четырехбалльной системе:

«отлично»;

«хорошо»;

«удовлетворительно»;

«неудовлетворительно».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Отлично	Хорошо	Удовл.	Неудовл.
ПК-5	знать методы и средства системного анализа, оптимизации, управления, принятия решений и обработки информации в машиностроении	Письменные вопросы	Ответ на все вопросы и дополнительные	Ответ на все вопросы	Ответ на 1 вопрос	Нет ответов
	уметь разрабатывать методы проектирования и анализа алгоритмов, используемых в производственных процессах	Письменные вопросы	Ответ на все вопросы и дополнительные	Ответ на все вопросы	Ответ на 1 вопрос	Нет ответов
	владеть современными методами исследования и информационно-коммуникационными технологиями	Письменные вопросы	Ответ на все вопросы и дополнительные	Ответ на все вопросы	Ответ на 1 вопрос	Нет ответов
ПК-3	знать математические методы обработки данных	Письменные вопросы	Ответ на все вопросы и дополнительные	Ответ на все вопросы	Ответ на 1 вопрос	Нет ответов
	уметь выполнять исследования процессов создания, накопления и обработки информации, включая анализ и создание моделей данных и знаний, языков их описания и манипулирования	Письменные вопросы	Ответ на все вопросы и дополнительные	Ответ на все вопросы	Ответ на 1 вопрос	Нет ответов
	владеть новыми методами исследования и обработки данных и их применению в самостоятельной научно-исследовательской деятельности в области профессиональной деятельности	Письменные вопросы	Ответ на все вопросы и дополнительные	Ответ на все вопросы	Ответ на 1 вопрос	Нет ответов

7.2 Примерный перечень оценочных средств (типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности)

7.2.1 Примерный перечень заданий для подготовки к тестированию

Не предусмотрено учебным планом

7.2.2 Примерный перечень заданий для решения стандартных задач

Не предусмотрено учебным планом

7.2.3 Примерный перечень заданий для решения прикладных задач

Не предусмотрено учебным планом

7.2.4 Примерный перечень вопросов для подготовки к зачету

Не предусмотрено учебным планом

7.2.5 Примерный перечень заданий для решения прикладных задач

1. Показатели и объекты (измерения). Интервальные данные.
2. Нечисловые данные.
3. Простейшие статистические характеристики.
4. Приведение к нормальной форме.
5. Оцифровка нечисловых данных.
6. Роль и сущность многомерных статистических методов.
7. Многомерное нормальное распределение как основная модель современных многомерных статистических методов.
8. Практическое применение многомерных методов.
9. Методы статистического оценивания многомерных параметров и проверки гипотез.
10. Особенности анализа количественных и качественных признаков. Методы шкалирования.
11. Большие данные. Терминология, история появления.
12. Технические сложности работы с большими данными.
13. Распределенная файловая система HDFS.
14. Базовая идея модели MapReduce, пример использования MapReduce.
15. Модель вычислений MapReduce. Реализация
16. MapReduce в Hadoop.
17. Основные понятия интеллектуального анализа данных.
18. Математическое обеспечение анализа данных.
19. Машинное обучение. Основные положения.
20. Библиотеки визуального анализа данных.
21. Задачи регрессии и классификации.
22. Линейная регрессия.
23. Логистическая регрессия.
24. Метрики оценивания ошибки обучения алгоритмов.
25. Исследование реальных данных.
26. Методы поиска выбросов в данных. Методы восстановления пропусков в данных. Работа с несбалансированными выборками.
27. Деревья решений. Методы построения деревьев.
28. Композиции алгоритмов. Разложение ошибки на смещение и разброс.
29. Случайный лес, его особенности.

7.2.6. Методика выставления оценки при проведении промежуточной аттестации

Экзамен проводится в письменной форме. На экзамен выдается 2 вопроса.

Оценка «отлично» - правильный ответ на 2 вопроса и дополнительный.

Оценка «хорошо» - правильный ответ на 2 вопроса.

Оценка «удовлетворительно» - правильный ответ на 1 вопрос.

Оценка «не удовлетворительно» - нет правильных ответов.

7.2.7 Паспорт оценочных материалов

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции	Наименование оценочного средства
1	Введение, основные понятия анализа данных	ПК-5, ПК-3	Контрольная работа, защита лабораторных работ
2	Математические объекты и методы в обработке больших данных	ПК-5, ПК-3	Контрольная работа, защита лабораторных работ
3	Визуальный анализ больших данных	ПК-5, ПК-3	Контрольная работа, защита лабораторных работ
4	Линейная регрессия и классификация	ПК-5, ПК-3	Контрольная работа, защита лабораторных работ
5	Оценивание качества алгоритмов	ПК-5, ПК-3	Контрольная работа, защита лабораторных работ
6	Особенности реальных больших данных	ПК-5, ПК-3	Контрольная работа, защита лабораторных работ

7.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Тестирование осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных тест-заданий на бумажном носителе. Время тестирования 30 мин. Затем осуществляется проверка теста экзаменатором и выставляется оценка согласно методики выставления оценки при проведении промежуточной аттестации.

Решение стандартных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

Решение прикладных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики

выставления оценки при проведении промежуточной аттестации. Защита курсовой работы, курсового проекта или отчета по всем видам практик осуществляется согласно требованиям, предъявляемым к работе, описанным в методических материалах. Примерное время защиты на одного студента составляет 20 мин.

8 УЧЕБНО МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ)

8.1 Перечень учебной литературы, необходимой для освоения дисциплины

1. Статистическое управление процессами. «Большие данные» : учебное пособие Ю. П. Адлер, Е. А. Черных. — Москва : Издательский Дом МИСиС, 2016. — 52 с. — ISBN 978-5-87623-969-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/64199.html>
2. Big Data. Методы и средства анализа : учебное пособие Л. И. Воронова, В. И. Воронов. — Москва : Московский технический университет связи и информатики, 2016. — 33 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/61463.html>
3. Анализ данных. Часть 2. Инструменты Data Mining : учебное пособие Ф. О. Федин, Ф. Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 308 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/26445.html>
4. Случайные данные. Структура и анализ В. И. Хищенко. — Москва : Техносфера, 2017. — 424 с. — ISBN 978-5-94836-497-1. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/84706.html>

8.2 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного программного обеспечения, ресурсов информационно-телекоммуникационной сети «Интернет», современных профессиональных баз данных и информационных справочных систем:

Лицензионное ПО:

- Windows Professional 8.1 (7 и 8) Single Upgrade MVL A Each Academic (многопользовательская лицензия)
- Microsoft Office Word 2007
- Microsoft Office Power Point 2007

Свободно распространяемое ПО:

- Adobe Acrobat Reader
- Python

Отечественное ПО:

- Яндекс.Браузер
- Архиватор 7z

Ресурс информационно-телекоммуникационной сети «Интернет»:

- Образовательный портал ВГТУ
- <http://www.edu.ru/>

Информационно-справочные системы:

- <http://window.edu.ru>
- <https://wiki.cchgeu.ru/>

Современные профессиональные базы данных:

<https://proglib.io>

9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Для проведения лекционных занятий необходима аудитория, оснащенная оборудованием для лекционных демонстраций и проекционной аппаратурой. Для проведения лабораторных работ необходима лаборатория с ПК, оснащенными программами для проведения лабораторного практикума и обеспечивающими возможность доступа к локальной сети кафедры и Интернет, из следующего перечня:

- 202/2
- 208/2
- 213/2

Аудитории располагаются по адресу: г.Воронеж, ул.Плехановская, 11

10. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

По дисциплине «Методы обработки больших данных» читаются лекции, проводятся лабораторные работы, выполняется курсовая работа.

Основой изучения дисциплины являются лекции, на которых излагаются наиболее существенные и трудные вопросы, а также вопросы, не нашедшие отражения в учебной литературе.

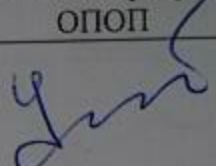
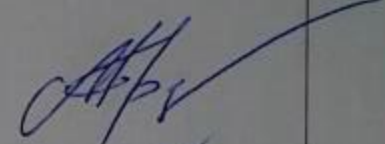
Лабораторные работы выполняются на лабораторном оборудовании в соответствии с методиками, приведенными в указаниях к выполнению работ. Выполнять этапы курсовой работы должны своевременно и в установленные сроки.

Контроль усвоения материала дисциплины производится проверкой курсовой работы, защитой курсовой работы.

Вид учебных занятий	Деятельность студента
---------------------	-----------------------

Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; пометать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначение вопросов, терминов, материала, которые вызывают трудности, поиск ответов в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на лекции или на практическом занятии.
Лабораторная работа	Лабораторные работы позволяют научиться применять теоретические знания, полученные на лекции при решении конкретных задач. Чтобы наиболее рационально и полно использовать все возможности лабораторных для подготовки к ним необходимо: следует разобрать лекцию по соответствующей теме, ознакомиться с соответствующим разделом учебника, проработать дополнительную литературу и источники, решить задачи и выполнить другие письменные задания.
Самостоятельная работа	Самостоятельная работа студентов способствует глубокому усвоению учебного материала и развитию навыков самообразования. Самостоятельная работа предполагает следующие составляющие: <ul style="list-style-type: none"> - работа с текстами: учебниками, справочниками, дополнительной литературой, а также проработка конспектов лекций; - выполнение домашних заданий и расчетов; - работа над темами для самостоятельного изучения; - участие в работе студенческих научных конференций, олимпиад; - подготовка к промежуточной аттестации.
Подготовка к промежуточной аттестации	Готовиться к промежуточной аттестации следует систематически, в течение всего семестра. Интенсивная подготовка должна начаться не позднее, чем за месяц-полтора до промежуточной аттестации. Данные перед экзаменом, экзаменом три дня эффективнее всего использовать для повторения и систематизации материала.

Лист регистрации изменений

№ п/п	Перечень вносимых изменений	Дата внесения изменений	Подпись заведующего кафедрой, ответственного за реализацию программы ОПОП
1	<p>Актуализация на основании Приказов Министерства науки и высшего образования Российской Федерации от 26.11.2020 г. №1456 «о внесении изменений в федеральные государственные образовательные стандарты высшего образования - бакалавриат по направлениям подготовки»</p>	31.08.2021	 
2	<p>Актуализация раздела 8.1 Перечень учебной литературы, необходимой для освоения дисциплины. Актуализирован раздел 8.2 в части состава используемого программного обеспечения, современных профессиональных баз данных и справочных информационных систем.</p>	31.08.2022	