Межгосударственное образовательное учреждение высшего образования

«Белорусско-Российский университет»

|  |
| --- |
| УТВЕРЖДАЮ |
| Первый проректор Белорусско-Российского университета |
| \_\_\_\_\_\_\_\_\_\_ Ю.В. Машин |
| «\_\_\_»\_\_\_\_\_\_\_ 2021 г. |
| Регистрационный № УД-\_\_\_\_\_\_\_\_\_/р |

**Методы анализа больших данных**

(наименование дисциплины)

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Направление подготовки** 01.03.04 Прикладная математика

**Направленность (профиль)** Разработка программного обеспечения

**Квалификация** Бакалавр

|  |  |
| --- | --- |
|  | **Форма обучения** |
| **Очная**  |
| Курс  |  4 |
| Семестр  | 7 |
| Лекции, часы | 60 |
| Лабораторные занятия, часы | 60 |
| Курсовая работа, семестр | 7 |
| Экзамен, семестр | 7 |
| Контактная работа по учебным занятиям, часы | 120 |
| Самостоятельная работа, часы | 168 |
| Всего часов / зачетных единиц | 288/8 |

Кафедра-разработчик программы: «Высшая математика»

Составители: Д.В. Роголев, канд. физ.-мат. наук; А.А. Романенко, канд. физ.-мат. наук, доц.; Т.Ю. Орлова, ст. пр.

(И.О. Фамилия, ученая степень, ученое звание)

Могилев, 2021

Рабочая программа составлена в соответствии с федеральным государственным образовательным стандартом высшего образования - бакалавриат по направлению подготовки 01.03.04 Прикладная математика № 11 от 10.01.2018 г., учебным планом рег. № 010304-2 от 26.03.2021 г.

Рассмотрена и рекомендована к утверждению кафедрой «Высшая математика»

27.05.2021 г., протокол № 9.

Зав. кафедрой \_\_\_\_\_\_\_\_\_\_\_\_\_\_ В.Г. Замураев

Одобрена и рекомендована к утверждению Научно-методическим советом

Белорусско-Российского университета

«16» июня 2021 г., протокол № 7.

Зам. председателя

Научно-методического совета \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ С.А. Сухоцкий

Рецензент:

Владимир Антонович Юревич, профессор кафедры техносферной безопасности и общей физики Белорусского государственного университета пищевых и химических технологий, доктор физико-математических наук, профессор

 (И.О. Фамилия, должность, ученая степень, ученое звание рецензента)

Рабочая программа согласована:

Ведущий библиотекарь \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Начальник учебно-методического

отдела \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ В.А. Кемова

**1 ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

**1.1 Цель учебной дисциплины**

Целями преподавания дисциплины являются:

• знакомство с основными методами анализа больших данных;

• формирование представления о типах задач, возникающих в области интеллектуального анализа данных и методах их решения, которые помогут студентам выявлять, формализовать и успешно решать практические задачи анализа данных, возникающие в процессе их профессиональной деятельности;

• приобретение навыков самостоятельного изучения отдельных тем дисциплины и решения типовых задач;

• усвоение полученных знаний студентами, а также формирование у них мотивации к самообразованию за счет активизации самостоятельной познавательной деятельности;

• сформировать интерес к математическим дисциплинам;

• показать историческую преемственность математических знаний.

**1.2 Планируемые результаты изучения дисциплины**

В результате освоения учебной дисциплины студент должен

**знать**:

- основные задачи математической статистики;

- основные методы и системы обработки больших данных;

- условия их применения и практические ограничения;

- современные методы и средства обработки больших данных;

- классификацию и суть математических моделей и методов, применяемых при обработке экспериментальных данных;

**уметь**:

- проводить предварительную обработку больших данных;

- решать задачи регрессионного, дисперсионного анализа;

- выбирать эффективные модели и методы для решения прикладных задач;

- выбирать методы решения задачи; **владеть**:

- технологиями и методиками сбора, предварительной подготовки и анализа экспериментальных данных;

- навыками разработки и отладки программ.

**1.3 Место учебной дисциплины в системе подготовки студента**

Дисциплина относится к Блоку 1 "Дисциплины (модули)" (Часть Блока 1, формируемая участниками образовательных отношений).

Перечень учебных дисциплин, изучаемых ранее, усвоение которых необходимо для изучения данной дисциплины:

- дискретная математика;

- линейная алгебра;

- математический анализ;

- программирование;

- аналитическая геометрия;

- математическая логика и теория алгоритмов;

- современные математические системы;

- теория вероятностей и случайные процессы;

- математическая статистика;

- объектно-ориентированное программирование;

- математическое программирование;

- случайные процессы.

Перечень учебных дисциплин, которые будут опираться на данную дисциплину:

- искусственный интеллект, машинное обучение, нейронные сети.

Кроме того, знания, полученные при изучении дисциплины на лабораторных занятиях будут применены при прохождении учебной и производственной практики, а также при подготовке выпускной квалификационной работы и дальнейшей профессиональной деятельности.

**1.4 Требования к освоению учебной дисциплины**

Освоение данной учебной дисциплины должно обеспечивать формирование следующих компетенций:

|  |  |
| --- | --- |
| Коды формируемых компетенций | Наименования формируемых компетенций |
| ПК-1 | Способен проводить научно-исследовательские разработки при исследовании самостоятельных тем |

**2 Структура и содержание дисциплины**

Вклад дисциплины в формирование результатов обучения выпускника (компетенций) и достижение обобщенных результатов обучения происходит путём освоения содержания обучения и достижения частных результатов обучения, описанных в данном разделе.

**2.1 Содержание учебной дисциплины**

|  |  |  |  |
| --- | --- | --- | --- |
| Номер тем | Наименование тем | Содержание | Коды формируемых компетенций |
| 1 | Введение в анализ больших данных. Обзор источников информации | История появления термина "большие данные". Три "V" (4, 5, 7) и три принципа работы с большими данными. Основные определения, термины, задачи анализа больших данных. Вопросы безопасности. Понятие Data Mining. Когнитивный анализ данных.Обзор источников информации для Big Data (открытые источники информации: статистические сборники, опубликованные отчёты и результаты исследований; доступ к закрытой информации). Методики сбора данных. | ПК-1 |
| 2 | Технологии хранения и обработки больших данных | Обзор технологий хранения больших данных. Базы данных. Системы управления базами данных. Модели данных. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных. | ПК-1 |
| 3 | Обзор методов анализа больших данных | Методы класса [Data Mining](https://ru.wikipedia.org/wiki/Data_Mining%22%20%5Co%20%22Data%20Mining), [кластерный анализ](https://ru.wikipedia.org/wiki/%D0%9A%D0%BB%D0%B0%D1%81%D1%82%D0%B5%D1%80%D0%BD%D1%8B%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7); [краудсорсинг](https://ru.wikipedia.org/wiki/%D0%9A%D1%80%D0%B0%D1%83%D0%B4%D1%81%D0%BE%D1%80%D1%81%D0%B8%D0%BD%D0%B3) ; [машинное обучение](https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5), включая [обучение с учителем](https://ru.wikipedia.org/wiki/%D0%9E%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_%D1%81_%D1%83%D1%87%D0%B8%D1%82%D0%B5%D0%BB%D0%B5%D0%BC) и [без учителя](https://ru.wikipedia.org/wiki/%D0%9E%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B1%D0%B5%D0%B7_%D1%83%D1%87%D0%B8%D1%82%D0%B5%D0%BB%D1%8F); [искусственные нейронные сети](https://ru.wikipedia.org/wiki/%D0%98%D1%81%D0%BA%D1%83%D1%81%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BD%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D1%81%D0%B5%D1%82%D1%8C), [сетевой анализ](https://ru.wikipedia.org/wiki/%D0%A1%D0%B5%D1%82%D0%B5%D0%B2%D0%BE%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7), [оптимизация](https://ru.wikipedia.org/wiki/%D0%9E%D0%BF%D1%82%D0%B8%D0%BC%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F_%28%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29); [распознавание образов](https://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D0%BE%D1%80%D0%B8%D1%8F_%D1%80%D0%B0%D1%81%D0%BF%D0%BE%D0%B7%D0%BD%D0%B0%D0%B2%D0%B0%D0%BD%D0%B8%D1%8F_%D0%BE%D0%B1%D1%80%D0%B0%D0%B7%D0%BE%D0%B2)*;* [прогнозная аналитика](https://ru.wikipedia.org/wiki/%D0%9F%D1%80%D0%BE%D0%B3%D0%BD%D0%BE%D0%B7%D0%BD%D0%B0%D1%8F_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D1%82%D0%B8%D0%BA%D0%B0); [имитационное моделирование](https://ru.wikipedia.org/wiki/%D0%98%D0%BC%D0%B8%D1%82%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5); [пространственный анализ](https://ru.wikipedia.org/w/index.php?title=%D0%9F%D1%80%D0%BE%D1%81%D1%82%D1%80%D0%B0%D0%BD%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85&action=edit&redlink=1) ; [статистический анализ](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7)  и [анализ временных рядов](https://ru.wikipedia.org/wiki/%D0%92%D1%80%D0%B5%D0%BC%D0%B5%D0%BD%D0%BD%D0%BE%D0%B9_%D1%80%D1%8F%D0%B4); визуализация аналитических данных . | ПК-1 |
| 4 | Основы анализа данных | Анализ данных в Microsoft Excel. Описательная статистика. Корреляционный анализ. Регрессионный анализ | ПК-1 |
| 5 | Методы классификации и прогнозирования. Деревья решений | Преимущества деревьев решений. Процесс конструирования дерева решений. Алгоритмы. | ПК-1 |
| 6 | Случайные леса | Мудрость толпы. Пример: предсказание криминальной активности. Ансамбли. Бэггинг. Ограничения. | ПК-1 |
| 7 | Метод опорных векторов. Метод "ближайшего соседа". Байесовская классификация | Метод опорных векторов. Линейный SVM. Метод "ближайшего соседа" или системы рассуждений на основе аналогичных случаев. Решение задачи классификации новых объектов. Решение задачи прогнозирования. Оценка параметра k методом кросс-проверки. Байесовская классификация. Байесовская фильтрация по словам. | ПК-1 |
| 8 | Краудсорсинг | Понятие, классические и современные виды краудсорсинга, сферы применения, примеры. | ПК-1 |
| 9 | Методы кластерного анализа. Иерархические методы  | Методы кластерного анализа. Меры сходства. Методы объединения или связи. Иерархический кластерный анализ в SPSS. | ПК-1 |
| 10 | Методы кластерного анализа. Итеративные методы. | Алгоритм k-средних. Алгоритм PAM (partitioning around Medoids). Предварительное сокращение размерности. Факторный анализ. Итеративная кластеризация в SPSS. Сравнительный анализ иерархических и неиерархических методов кластеризации. | ПК-1 |
| 11 | Алгоритмы и некоторые модификации алгоритмов кластерного анализа | Алгоритмы BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies), WaveCluster, CLARA (Clustering LARge Applications), Clarans, CURE, DBScan. | ПК-1 |
| 12 | Теория распознавания образов | Постановка задачи, методы распознавания образов, Персептрон. Примеры. | ПК-1 |
| 13 | Прогнозная аналитика | Основные типы и методы. Сфера применения. | ПК-1 |
| 14 | Имитационное моделирование | Понятие, виды имитационного моделирования. Реализация. Области применения. | ПК-1 |
| 15 | Смешение и интеграция данных | Уровни интеграции. Семантическая интеграция. Виртуализация данных. | ПК-1 |
| 16 | Статистический анализ | Обзор статистических методов. Одномерные статистические методы. | ПК-1 |
| 17 | Статистический анализ | Многомерные статистические методы. | ПК-1 |
| 18 | Визуализация аналитических данных | Основные правила, приёмы и инструменты визуализации. | ПК-1 |
| 19 | Понятие Data Mining | Сравнение статистики, машинного обучения и Data Mining. Развитие технологии баз данных. Отличия Data Mining от других методов анализа данных. Перспективы технологии Data Mining. | ПК-1 |
| 20 | Методы и стадии Data Mining | Классификация стадий Data Mining. Классификация методов Data Mining: статистические, кибернетические методы. Свойства методов Data Mining. | ПК-1 |
| 21 | Комплексный подход к внедрению Data Mining | Классификация СППР. | ПК-1 |
| 22 | OLAP и хранилища данных в СППР | OLAP-системы, OLAP-продукты. Интеграция OLAP и Data Mining. Хранилища данных. Преимущества использования хранилищ данных. | ПК-1 |
| 23 | Процесс Data Mining. Начальные этапы | Анализ предметной области, постановка задачи, подготовка данных. | ПК-1 |
| 24 | Очистка данных | Инструменты очистки данных. Выводы по подготовке данных. | ПК-1 |
| 25 | Построение и использование модели | Моделирование. Виды моделей. Математическая модель: построение, проверка и оценка моделей, выбор, применение, коррекция и обновление модели. Погрешности в процессе Data Mining. | ПК-1 |
| 26 | Рынок инструментов Data Mining | Поставщики Data Mining. Классификация инструментов Data Mining. Программное обеспечение Data Mining для поиска ассоциативных правил. Программное обеспечение для решения задач кластеризации и сегментации. Программное обеспечение для решения задач классификации. Программное обеспечение Data Mining для решения задач оценивания и прогнозирования. | ПК-1 |
| 27 | Современные программные средства анализа больших объёмов информации | Обзор современных популярных программных средств анализа данных: Statistica, SPSS, Excel, R-Studio, R и другие; их преимущества и недостатки. | ПК-1 |

**2.2 Учебно-методическая карта учебной дисциплины**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| № недели | Лекции(наименование тем) | Часы | Лабораторные занятия | Часы | Самостоятельная работа, часы | Форма контроля знаний | Баллы (max) |
| Модуль 1 |
| 1 | 1. Введение в анализ больших данных. Обзор источников информации | 2 | 1. Предварительная обработка значений временных рядов. | 2 | 2 |  |  |
|  | 2. Технологии хранения и обработки больших данных | 2 | 1. Предварительная обработка значений временных рядов. | 2 | 4 | ЗЛР | 6 |
| 2 | 3. Обзор методов анализа больших данных | 2 | 2. Корреляционный и регрессионный анализ. | 2 | 2 |  |  |
|  | 4. Основы анализа данных | 2 | 2. Корреляционный и регрессионный анализ. | 2 | 4 |  |  |
| 3 | 5. Методы классификации и прогнозирования. Деревья решений | 2 | 2. Корреляционный и регрессионный анализ. | 2 | 4 | ЗЛР | 6 |
|  | 6. Случайные леса | 2 | 3. Дискриминантный анализ. | 2 | 2 |  |  |
| 4 | 7. Метод опорных векторов. Метод "ближайшего соседа". Байесовская классификация | 2 | 3. Дискриминантный анализ. | 2 | 4 |  |  |
|  | 8. Краудсорсинг | 2 | 3. Дискриминантный анализ. | 2 | 2 | ЗЛР | 6 |
| 5 | 9. Методы кластерного анализа. Иерархические методы | 2 | 4. Метод главных компонент. | 2 | 4 |  |  |
|  | 9. Методы кластерного анализа. Иерархические методы | 2 | 4. Метод главных компонент. | 2 | 2 | ЗЛР | 6 |
| 6 | 10. Методы кластерного анализа. Итеративные методы. | 2 | 5. Кластерный анализ. Иерархические методы. | 2 | 4 |  |  |
|  | 10. Методы кластерного анализа. Итеративные методы. | 2 | 5. Кластерный анализ. Иерархические методы. | 2 | 4 |  |  |
| 7 | 11. Алгоритмы и некоторые модификации алгоритмов кластерного анализа | 2 | 5. Кластерный анализ. Иерархические методы. | 2 | 4 | ЗЛР | 6 |
|  | 12. Теория распознавания образов | 2 | 6. Неиерархические методы кластерного анализа. | 2 | 2 |  |  |
| 8 | 13. Прогнозная аналитика | 2 | 6. Неиерархические методы кластерного анализа. | 2 | 2 |  |  |
|  | 14. Имитационное моделирование | 2 | 6. Неиерархические методы кластерного анализа. | 2 | 2 | ПКУ | 30 |
| Модуль 2 |
| 9 | 15. Смешение и интеграция данных | 2 | 7. Компонентный анализ. | 2 | 2 | ЗЛР | 6 |
|  | 16. Статистический анализ | 2 | 7. Компонентный анализ. | 2 | 4 |  |  |
| 10 | 17. Статистический анализ | 2 | 7. Компонентный анализ. | 2 | 4 | ЗЛР | 6 |
|  | 18. Визуализация аналитических данных | 2 | 8. Факторный анализ. | 2 | 4 |  |  |
| 11 | 19. Понятие Data Mining | 2 | 8. Факторный анализ. | 2 | 2 |  |  |
|  | 20. Методы и стадии Data Mining | 2 | 8. Факторный анализ. | 2 | 2 | ЗЛР | 6 |
| 12 | 21. Комплексный подход к внедрению Data Mining | 2 | 9. Многомерный статистический анализ. | 2 | 4 |  |  |
|  | 22. OLAP и хранилища данных в СППР | 2 | 9. Многомерный статистический анализ. | 2 | 4 |  |  |
| 13 | 23. Процесс Data Mining. Начальные этапы | 2 | 9. Многомерный статистический анализ. | 2 | 4 | ЗЛР | 6 |
|  | 24. Очистка данных | 2 | 10. Анализ данных с помощью технологии Data Mining. | 2 | 4 |  |  |
| 14 | 25. Построение и использование модели | 2 | 10. Анализ данных с помощью технологии Data Mining. | 2 | 4 |  |  |
|  | 26. Рынок инструментов Data Mining | 2 | 10. Анализ данных с помощью технологии Data Mining. | 2 | 2 |  |  |
| 15 | 27. Современные программные средства анализа больших объёмов информации | 2 | 10. Анализ данных с помощью технологии Data Mining. | 2 | 4 | ЗЛР | 6 |
|  | 27. Современные программные средства анализа больших объёмов информации | 2 | 10. Анализ данных с помощью технологии Data Mining. | 2 | 4 | ПКУ | 30 |
| 1-15 | Выполнение курсовой работы |  |  |  | 36 |  |  |
| 16-18 |  |  |  |  | 36 | ПА(экзамен) | 40 |
|  | Итого | 60 |  | 60 | 168 |  | 100 |

Принятые обозначения:

*Текущий контроль* –

ЗЛР – защита лабораторной работы;

ПКУ – промежуточный контроль успеваемости;

ПА – промежуточная аттестация.

Итоговая оценка определяется как сумма текущего контроля и промежуточной аттестации и соответствует баллам:

Экзамен

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Оценка | Отлично | Хорошо | Удовлетворительно | Неудовлетворительно |
| Баллы | 87-100 | 65-86 | 51-64 | 0-50 |

**2.3 Требования к курсовой работе**

Целью курсовой работы является то, что её выполнение студентом способствует углублению знаний и умений, полученных им в ходе теоретических и лабораторных занятий, прививает навыки самостоятельного изучения материала по теме курсовой работы, а также развивает компетенции аналитической, исследовательской и проектной деятельности, работы с информацией.

Примерная тематика курсовых работ представлена в приложении хранится на кафедре.

Курсовая работа включает теоретическую и практическую части общим объёмом 25-30 страниц.

Перечень этапов выполнения курсовой работы и количества баллов за каждый из них представлен в таблице.

|  |  |  |
| --- | --- | --- |
| Этап выполнения | Минимум | Максимум |
| Теоретическая часть | 18 | 30 |
| Практические исследования | 15 | 25 |
| Оформление работы | 3 | 5 |
| **Итого за выполнение курсовой работы** | **36** | **60** |
| **Защита курсовой работы** | **15** | **40** |

Итоговая оценка курсовой работы представляет собой сумму баллов за её выполнение и защиту и выставляется в соответствии со шкалой:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Оценка | Отлично | Хорошо | Удовлетворительно | Неудовлетворительно |
| Баллы | 87-100 | 65-86 | 51-64 | 0-50 |

**3 ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ**

При изучении дисциплины используется модульно-рейтинговая система оценки знаний студентов. Применение форм и методов проведения занятий при изучении различных тем курса представлено в таблице.

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Форма проведения занятия** | **Вид аудиторных занятий** | **Всего часов** |
| **Лекции** | **Лабораторные занятия** |
| 1 | Традиционные | 1, 2, 5-26 |  | 52 |
| 2 | Мультимедиа | 3, 4, 27 |  | 8 |
| 3 | С использованием ЭВМ |  | 1-10 | 60 |
|  | **ИТОГО** |  |  | 120 |

**4 ОЦЕНОЧНЫЕ СРЕДСТВА**

Используемые оценочные средства по учебной дисциплине представлены в таблице и хранятся на кафедре.

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Вид оценочных средств** | **Количество комплектов** |
| 1 | Вопросы к экзамену | 1 |
| 2 | Экзаменационные билеты | 1 |
| 3 | Вопросы к защите лабораторных работ  | 1 |

**5 Методика и критерии оценки компетенций студентов**

**5.1 Уровни сформированности компетенций**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Уровни сформированности компетенции** | **Содержательное описание уровня** | **Результаты обучения** |
| *Компетенция* ПК-1 Способен проводить научно-исследовательские разработки при исследовании самостоятельных тем |
| *Код и наименование индикатора достижения компетенции*ПК-1.13 Способен применять методы анализа больших данных при проведении научно-исследовательских разработок |
| 1 | Пороговый уровень | Базовые знания в объеме рабочей программы (знание основных методов анализа), умение решать типовые задачи под руководством преподавателя. | Имеет представление о методах анализа больших данных при проведении научно-исследовательских разработок. |
| 2 | Продвинутый уровень | Полные знания в объеме рабочей программы, правильное использование терминологии, способность самостоятельно решать типовые задачи учебной дисциплины. | Умеет использовать методы анализа больших данных при проведении научно-исследовательских разработок. |
| 3 | Высокий уровень | Систематизированные, глубокие и полные знания в объеме рабочей программы, точное использование научной терминологии и свободное владение инструментарием учебной дисциплины, умение анализировать и применять теоретические знания при самостоятельном решении типовых учебных задач и задач повышенной сложности, способность делать обоснованные выводы. | Владеет навыками применения методов анализа больших данных при проведении научно-исследовательских разработок. |

**5.2 Методика оценки знаний, умений и навыков студентов**

|  |  |
| --- | --- |
| Результаты обучения | Оценочные средства |
| *Компетенция* ПК-1 |
| Пороговый уровень | Вопросы к экзамену.Экзаменационные билеты.Вопросы к защите лабораторных работ. |
| Продвинутый уровень | Вопросы к экзамену.Экзаменационные билеты.Вопросы к защите лабораторных работ. |
| Высокий уровень | Вопросы к экзамену.Экзаменационные билеты.Вопросы к защите лабораторных работ. |

**5.3 Критерии оценки лабораторных работ**

Лабораторные работы (ЗЛР) оцениваются до 6 баллов:

**0-1 баллов** - оборудование и методы использованы неправильно, проявлена неудовлетворительная теоретическая подготовка, необходимые навыки и умения не освоены, результат работы не соответствует её целям;

**2-3 баллов** - оборудование и методы частично использованы правильно, проявлена удовлетворительная теоретическая подготовка, необходимые навыки и умения частично освоены, результат работы частично соответствует её целям;

**4-5 баллов** - оборудование и методы в основном использованы правильно, проявлена хорошая теоретическая подготовка, необходимые навыки и умения в основном освоены, результат работы в основном соответствует её целям;

**6 баллов** - оборудование и методы использованы правильно, проявлена отличная теоретическая подготовка, необходимые навыки и умения полностью освоены, результат работы полностью соответствует её целям.

**5.4 Критерии оценки курсовой работы**

Итоговая оценка курсовой работы представляет собой сумму баллов за её выполнение и защиту и выставляется в соответствии со шкалой:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Оценка | Отлично | Хорошо | Удовлетворительно | Неудовлетворительно |
| Баллы | 87-100 | 65-86 | 51-64 | 0-50 |

Оценка «**отлично**» выставляется, если работа выполнена в соответствии с утвержденным планом, полностью раскрыто содержание каждого вопроса, студентом сформулированы собственные аргументированные выводы по теме работы. Оформление работы соответствует предъявляемым требованиям. При защите работы студент свободно владеет материалом и отвечает на вопросы.

Оценка «**хорошо**» выставляется, если работа выполнена в соответствии с утвержденным планом, полностью раскрыто содержание каждого вопроса. Незначительные замечания к оформлению работы. При защите работы студент владеет материалом, но отвечает не на все вопросы.

Оценка «**удовлетворительно**» выставляется, если работа выполнена в соответствии с утвержденным планом, но не полностью раскрыто содержание каждого вопроса. Студентом не сделаны собственные выводы по теме работы. Грубые недостатки в оформлении работы. При защите работы студент слабо владеет материалом, отвечает не на все вопросы.

Оценка «**неудовлетворительно**» выставляется, если работа выполнена не в соответствии с утвержденным планом, не раскрыто содержание каждого вопроса. Студентом не сделаны выводы по теме работы. Грубые недостатки в оформлении работы. При защите работы студент не владеет материалом, не отвечает на вопросы.

**5.6 Критерии оценки экзамена**

Итоговая оценка на экзамене по пятибалльной системе определяется как сумма баллов промежуточного контроля успеваемости и текущей аттестации (экзамена) и соответствует суммарным баллам:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Оценка | Отлично | Хорошо | Удовлетворительно | Неудовлетворительно |
| Баллы | 87-100 | 65-86 | 51-64 | 0-50 |

При этом промежуточный контроль успеваемости оценивается до 60 баллов, а текущая аттестация (экзамен) – до 40 баллов. Экзаменационный билет состоит из 4 вопросов (3 теоретических вопроса и 1 задача), за каждое задание можно набрать до 10 баллов.

Оценка «**отлично**» выставляется за: систематизированные, глубокие и полные знания в объёме рабочей программы, точное использование научной терминологии и свободное владение инструментарием учебной дисциплины, умение анализировать и применять теоретические знания при самостоятельном решении типовых учебных задач и задач повышенной сложности, способность делать обоснованные выводы.

Оценка «**хорошо**» выставляется за: полные знания в объёме рабочей программы, правильное использование терминологии, способность самостоятельно решать типовые задачи учебной дисциплины.

Оценка «**удовлетворительно**» выставляется за: обладание базовыми знаниями (владеет терминологией, знает определение понятий) в объёме рабочей программы, достаточными для усвоения последующих дисциплин, умение решать простейшие типовые задачи.

Оценка «**неудовлетворительно**» выставляется за: фрагментарные знания по базовым вопросам в объёме рабочей программы, недостаточными для усвоения последующих дисциплин, неуверенное использование терминологии, неумение решать типовые задачи.

**6 МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ И ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ ПО УЧЕБНОЙ ДИСЦИПЛИНЕ**

Самостоятельная работа студентов (СРС) направлена на закрепление и углубление освоения учебного материала, развитие практических умений. СРС включает следующие виды самостоятельной работы студентов:

- изучение основной и дополнительной литературы;

- решение индивидуальных задач во время проведения лабораторных занятий под контролем преподавателя.

Перечень контрольных вопросов и заданий для самостоятельной работы студентов приведен в приложении и хранится на кафедре.

Для СРС рекомендуется использовать источники, приведенные в п. 7.

**7 УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ**

**7.1 Основная литература**

|  |  |  |  |
| --- | --- | --- | --- |
| № | Автор, название, место издания, издательство, год издания учебной литературы | Гриф | Кол-воэкз. |
| 1 | Лонг, Д. R. Книга рецептов: проверенные рецепты для статистики, анализа и визуализации данных: практическое руководство [Электронный ресурс] / Д. Лонг, П. Титор ; пер. с англ. Д. А. Беликова. - Москва: ДМК Пресс, 2020. — 510 с.— (Высшее образование: Бакалавриат). — **Режим доступа:** **http://znanium.com/** | — | ЭБС "Znanium" |
| 2 | Безруков А.И., Алексенцева О.Н.  Математическое и имитационное моделирование [Электронный ресурс]: учеб. пособие. — Москва: ИНФРА-М, 2019. — 227 с. — (Высшее образование: Бакалавриат). — **Режим доступа:** **http://znanium.com/** | — | ЭБС "Znanium" |
| 3 | Шорохова И.С., Кисляк Н.В., Мариев О.С. Статистические методы анализа [Электронный ресурс]: учеб. пособие .  — М.: Флинта, 2017. - 300 с. — (Высшее образование: Бакалавриат). — **Режим доступа:** **http://znanium.com/** | — | ЭБС "Znanium" |

**7.2 Дополнительная литература**

| № | Автор, название, место издания, издательство, год издания учебной литературы | Гриф | Кол-во экз. |
| --- | --- | --- | --- |
| 1 | Дадян Э.Г. Данные: хранение и обработка [Электронный ресурс]: учебник. — М.: ИНФРА-М, 2019. — 205 с. — (Высшее образование: Бакалавриат). — **Режим доступа:** **http://znanium.com/** | — | ЭБС "Znanium" |
| 2 | Форман Д.; пер. с англ. Соколовой А. Много цифр. Анализ больших данных при помощи Excel [Электронный ресурс]: учебное пособие. — М.: Альпина Пабл., 2016. — 461 с. — **Режим доступа: http://znanium.com/** | — | ЭБС "Znanium" |

|  |  |  |  |
| --- | --- | --- | --- |
| 3 | Юре Лесковец, Ананд Раджараман, Джеффри Д. Ульман ; пер. с англ. А.А. Слинкина. Анализ больших наборов данных  [Электронный ресурс]: практическое пособие. — М.: ДМК Пресс, 2016. — 498 с. (Высшее образование: Бакалавриат). — **Режим доступа: http://znanium.com/** | — | ЭБС "Znanium" |
| 4 | Груздев А.В. Прогнозное моделирование в IBM SPSS Statistics и R: Метод деревьев решений [Электронный ресурс]: практическое руководство. — М.: ДМК Пресс, 2016. — 278 с. (Высшее образование: Бакалавриат). — **Режим доступа: http://znanium.com/** | — | ЭБС "Znanium" |
| 5 | Кобелев Н.Б., Половников В.А., Девятков В.В.; Под общ. ред. д-ра экон. наук Кобелева Н.Б. Имитационное моделирование [Электронный ресурс]: учебное пособие. — М.: КУРС: НИЦ Инфра-М, 2013. — 368 с. (Высшее образование: Бакалавриат). — **Режим доступа: http://znanium.com/** | — | ЭБС "Znanium" |

**7.3 Перечень ресурсов сети Интернет по изучаемой дисциплине**

http://znanium.com/

habr.com

github.com

**7.4 Перечень наглядных и других пособий, методических рекомендаций по проведению учебных занятий, а также методических материалов к используемым в образовательном процессе техническим средствам**

**7.4.1 Методические рекомендации**

1. Т.Ю. Орлова, Д.В. Роголев, А.А. Романенко. Методы анализа больших данных. Методические рекомендации к лабораторным работам для студентов направления подготовки01.03.04 "Прикладная математика" дневной формы обучения. 2021 г. (электронный вариант)

**7.4.2 Информационные технологии**

Тема 3 - Обзор методов анализа больших данных.

Тема 4 - Основы анализа данных.

Тема 27 - Современные программные средства анализа больших объёмов информации.

**7.4.3 Перечень программного обеспечения, используемого в образовательном процессе**

Свободно распространяемое ПО:

R-project

GNU Octave

Python

**8 МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ**

Материально-техническое обеспечение дисциплины содержится в паспорте лаборатории ауд. 405, рег. номер ПУЛ-4.535-405/1-20.

**Методы анализа больших данных**

(наименование дисциплины)

**АННОТАЦИЯ**

**К РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ**

**Направление подготовки** 01.03.04 Прикладная математика

**Направленность (профиль)** Разработка программного обеспечения

|  |  |
| --- | --- |
|  | **Форма обучения** |
| **Очная**  |
| Курс  |  4 |
| Семестр  | 7 |
| Лекции, часы | 60 |
| Лабораторные занятия, часы | 60 |
| Курсовая работа, семестр | 7 |
| Экзамен, семестр | 7 |
| Контактная работа по учебным занятиям, часы  | 120 |
| Самостоятельная работа, часы | 168 |
| Всего часов / зачетных единиц | 288/8 |

**1. Цель учебной дисциплины**

Целями преподавания дисциплины являются:

• знакомство с основными методами анализа больших данных;

• формирование представления о типах задач, возникающих в области интеллектуального анализа данных и методах их решения, которые помогут студентам выявлять, формализовать и успешно решать практические задачи анализа данных, возникающие в процессе их профессиональной деятельности;

• приобретение навыков самостоятельного изучения отдельных тем дисциплины и решения типовых задач;

• усвоение полученных знаний студентами, а также формирование у них мотивации к самообразованию за счет активизации самостоятельной познавательной деятельности;

• сформировать интерес к математическим дисциплинам;

• показать историческую преемственность математических знаний.

**2. Планируемые результаты изучения дисциплины**

В результате освоения учебной дисциплины студент должен

**знать**:

- основные задачи математической статистики;

- основные методы и системы обработки больших данных;

- условия их применения и практические ограничения;

- современные методы и средства обработки больших данных;

- классификацию и суть математических моделей и методов, применяемых при обработке экспериментальных данных;

**уметь**:

- проводить предварительную обработку больших данных;

- решать задачи регрессионного, дисперсионного анализа;

- выбирать эффективные модели и методы для решения прикладных задач;

- выбирать методы решения задачи; **владеть**:

- технологиями и методиками сбора, предварительной подготовки и анализа экспериментальных данных;

- навыками разработки и отладки программ.

**3. Требования к освоению учебной дисциплины**

Освоение данной учебной дисциплины должно обеспечивать формирование следующих компетенций:

|  |  |
| --- | --- |
| Коды формируемых компетенций | Наименования формируемых компетенций |
| ПК-1 | Способен проводить научно-исследовательские разработки при исследовании самостоятельных тем |

**4. Образовательные технологии**

Традиционные, мультимедиа, с использованием ЭВМ.