Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 03.11.2023 16:44:12
Уникальный программный ключ:
8db180d1a3f02ac9e60521a5672742735c18b1d6

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ**

Федеральное государственное автономное образовательное учреждение высшего образования

«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**УТВЕРЖДЕНО**
Декан факультета
Информационных технологий
_____ / Демидов Д.Г. /
«27» апреля 2022 г.

# РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

## «Методы работы с большими данными»

Направление подготовки
**09.03.01 «Информатика и вычислительная техника»**

Профиль подготовки
«**Киберфизические системы**»

Квалификация (степень) выпускника
**Бакалавр**

Форма обучения
**Очная**

Москва 2022 г.

## 1. Цели освоения дисциплины

К основным целям освоения дисциплины «Методы работы с большими данными» относится: 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.

К основным задачам дисциплины «Методы работы с большими данными» относятся:

1. Изучение основных методов решения на ЭВМ задач анализа и интерпретации данных, получаемых с помощью различного рода информационно-измерительных систем. Знание базовых алгоритмов анализа и интерпретации данных.
2. Формирование навыков применения методов решения на ЭВМ задач анализа и интерпретации данных при разработке алгоритмов анализа и обработки измерительной информации. Умение использовать стандартную терминологию, определения и обозначения.
3. Освоение основных тенденций развития теории и практики анализа и интерпретации данных. Приобретение практических навыков работы с современными пакетами прикладных программ для решения задач анализа и интерпретации данных.

## 2. Место дисциплины в структуре ОПОП

Дисциплина относится к числу учебных дисциплин формируемые участниками образовательных отношений части Дисциплины по выбору студента «Элективные дисциплины» основной профессиональной образовательной программы.

Дисциплина взаимосвязана логически и содержательно-методически со следующими дисциплинами и практиками ОПОП:

- Алгоритмическое программирование;
- Искусственные нейронные сети;
- Численные методы в задачах управления

## 3. Перечень планируемых результатов обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

В результате освоения дисциплины у обучающихся формируются следующие компетенции и должны быть достигнуты следующие результаты обучения как этап формирования соответствующих компетенций.

| Код компетенции | В результате освоения образовательной программы обучающийся должен обладать | Перечень планируемых результатов обучения по дисциплине |
|---|---|---|
| ПК-2 | **ПК-2.** Способен осуществлять концептуальное, функциональное и логическое проектирование систем среднего и крупного масштаба и сложности. | ПК-2.1. Знает:<br>методы целеполагания;<br>теорию ключевых показателей деятельности;<br>методы концептуального проектирования;<br>стандарты оформления технических заданий;<br>теорию тестирования;<br>методы оценки качества программных систем;<br>методы тестирования;<br>международные стандарты на структуру документов требований;<br>нормативные и методические материалы по созданию документов требований к системам.<br>ПК-2.2. Умеет:<br>формулировать цели, исходя из анализа проблем, потребностей и возможностей;<br>разрабатывать технико-экономическое обоснование;<br>декомпозировать функции на подфункции;<br>алгоритмизировать деятельность;<br>разрабатывать структуры типовых документов;<br>исполнять ручные тесты.<br>ПК-2.3. Владеет:<br>навыками логического мышления;<br>средствами автоматизации проектирования ПО. |

В процессе освоения образовательной программы данные компетенции, в том числе их отдельные компоненты, формируются поэтапно в ходе освоения обучающимися дисциплин (модулей), практик в соответствии с учебным планом и календарным графиком учебного процесса.

## 4. Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 4 зачетных единицы, 144 академических часа (из них 72 часа – самостоятельная работа студентов).

Разделы дисциплины изучаются на четвертом курсе в седьмом семестре, форма контроля - экзамен.

| № п/п | Вид учебной работы | Количество часов | Семестр |
|---|---|---|---|
| **1.** | **Аудиторные занятия** | **72** | 72 |
| | В том числе: | | |
| 1.1 | Лекции | 36 | 36 |
| 1.2 | Практические занятия | | |
| 1.3 | Лабораторные занятия | 36 | 36 |
| **2.** | **Самостоятельная работа** | **72** | **72** |
| **3** | **Промежуточная аттестация** | | |
| | Экзамен | + | + |
| | Итого: | **144** | 144 |

Содержание дисциплины.

**Введение**

Предмет дисциплины и ее задачи. Краткие сведения о развитии теории анализа и интерпретации данных. Структура, содержание дисциплины, ее связи с другими дисциплинами учебного плана специальности. Перечень дисциплин и их разделов, усвоение которых необходимо студентам для изучения курса «Анализ и интерпретация данных».

**Тема 1. Основные понятия дисциплины**

Введение в анализ данных. Проблема обработки данных. Матрица данных. Гипотезы компактности и скрытых факторов. Структура матрицы данных и задачи обработки. Матрица объект-объект и признак-признак. Расстояние и близость. Измерение признаков. Отношения и их представление. Основные проблемы измерений. Основные типы шкал. Проблема адекватности. Основные задачи анализа и интерпретации данных .

**Тема 2. Классификация данных с использованием детерминированных моделей**

Решающие поверхности и дискриминантные функции. Линейные дискриминантные функции. Линейная разделимость. Кусочно-линейные дискриминантные функции. Нелинейные дискриминантные функции. Фимашины. Потенциальные функции как дискриминантные функции. Пространство весов. Процедуры обучения с коррекцией ошибок: правило с фиксированным приращением, правило абсолютной коррекции, частично корректирующее правило. Обобщенные градиентные методы. Персептронный критерий. Процедуры обучения на основе минимальной среднеквадратичной ошибки: псевдоинверсный метод, метод Хо-Кашпа

**Тема 3. Классификация данных на основе статистических моделей**

Функция потерь. Байесовская дискриминантная функция. Принятие решение по максимуму правдоподобия. Оптимальная дискриминантная функция для нормально распределенных образов. Дискриминантная функция Фишера. Множественный дискриминантный анализ. Пошаговый дискриминантный анализ. Ошибки классификации. Примеры построения статистических дискриминантных функций для различных статистических моделей данных. Обучение для статистических дискриминантных функций. Оценки максимального правдоподобия, байесовские оценки. Непараметрическое оценивание. Парзеновские окна, метод непараметрического оценивания на основе К-ближайшего соседства

**Тема 4. Кластер-анализ**

Основные типы задач кластер-анализа. Меры подобия и функции расстояния Выбор критерия кластеризации. Кластерные методы, основанные на евклидовой метрике. Иерархическая кластеризация. Метод Квнутригрупповых средних. Использование методов теории графов в задачах кластеризации. Кластеризация на основе анализа плотностей вероятностей

**Тема 5. Методы снижения размерностей данных**

Анализ матриц исходных данных. Метод главных компонент. Корреляционная матрица и ее основные свойства. Собственные векторы и собственные числа корреляционной матрицы. Приведение корреляционной матрицы к диагональной форме. Геометрическая интерпретация главных компонент на плоскости. Модели факторного анализа. Оценка факторных нагрузок методом максимального правдоподобия и центроидным методом. Вращение факторов и их интерпретация. Использование кластеризации признаков для снижения размерности. Многомерное шкалирование. Метрический и неметрический подход к многомерному шкалированию. Методы ортогонального проектирования. Нелинейные методы многомерного шкалирования. Многомерное шкалирование неметрических данных. Многомерные развертки

**Тема 6. Системы DATA MINING в задачах анализа и интерпретации данных** (Понятие об интеллектуальных системах анализа и интерпретации данных. DATA MINING - системы извлечения новых знаний из данных. Типы систем DATA MINING -предметно-ориентированные аналитические системы, статистические пакеты, нейронные сети, деревья решений, обнаружение логических закономерностей, генетические алгоритмы, системы визуализации многомерных данных

**Тема 7. Современные пакеты прикладных программ для решения задач обработки экспериментальных данных**. Библитеки Sklearn, Seaborn, Keras, Tensorflow. Табличные процессоры и базы данных в задачах обработки данных. Виды статистических пакетов.

## 5. Образовательные технологии

Рекомендации для организации учебно-методического обеспечения самостоятельной работы по дисциплине.

**Самостоятельная работа студентов** - это индивидуальная или коллективная учебная деятельность, осуществляемая без непосредственного руководства преподавателя или совместно с ним.

Самостоятельная работа есть особо организованный вид учебной деятельности, проводимый с целью повышения эффективности подготовки студентов к последующим занятиям, формирования у них навыков самостоятельной отработки учебных заданий, а также овладения методикой организации своего самостоятельного труда в целом.

Она призвана обеспечить более глубокое, творческое усвоение понятийного аппарата изучаемой дисциплины, содержания основных нормативно-правовых актов и литературы по данному учебному курсу.

Организация самостоятельной работы студентов должна строиться по системе поэтапного усвоения материала.

Метод поэтапного изучения включает в себя предварительную подготовку, непосредственное изучение теоретического содержания источника, обобщение полученных знаний.

Предварительная подготовка включает в себя уяснение цели изучения материала, оценка широты информационной базы анализируемого вопроса, выяснение его научной и практической актуальности.

Изучение теоретического содержания заключается в выделении и уяснении ключевых понятий и положений, выявлении их взаимосвязи и систематизации.

Обобщение полученных знаний подразумевает широкое осмысление теоретических положений через определение их места в общей структуре изучаемой дисциплины и их значимости для практической деятельности.

Самостоятельная работа призвана, прежде всего, сформировать у студентов навыки работы с литературой.

При анализе литературных источников студенты должны научиться правильно фиксировать основные реквизиты материалов (полное

официальное название, автор, где публикован, когда опубликован). Следует обратить особое внимание на новую для студента терминологию, без знания которой он не сможет усвоить содержание материалов, а в дальнейшем и ключевых положений изучаемой дисциплины в целом. В этих целях, как показывает опыт, незаменимую помощь оказывают всевозможные справочные издания, прежде всего, энциклопедического характера.

При самостоятельном изучении основной рекомендованной литературы студентам необходимо обратить главное внимание на узловые положения, излагаемые в тексте.

Для этого - необходимо внимательно ознакомиться с содержанием соответствующего блока информации, структурировать его и выделить в нем центральное звено. Обычно это бывает ключевое определение или совокупность сущностных характеристик рассматриваемого объекта. Для того, чтобы убедиться насколько глубоко усвоено содержание темы, в конце соответствующих глав и параграфов учебных пособий обычно дается перечень контрольных вопросов, на которые студент должен уметь дать четкие и конкретные ответы.

Работа с дополнительной литературой предполагает умение студентов выделять в ней необходимый аспект изучаемой темы (то, что в данном труде относится непосредственно к изучаемой теме). Это важно в связи с тем, что к дополнительной литературе может быть отнесен широкий спектр текстов (учебных, научных, художественных, публицистических и т.д.), В которых исследуемый вопрос рассматривается либо частично, либо с какой-то одной точки зрения, порой нетрадиционной. В своей совокупности изучение таких подходов существенно обогащает научный кругозор студентов. В данном контексте следует учесть, что дополнительную литературу целесообразно прорабатывать, во-первых; на базе уже освоенной основной литературы, и, во-вторых, изучать комплексно, всесторонне, не абсолютизируя чью-либо субъективную точку зрения.

Обязательный элемент самостоятельной работы студентов с литературными источниками - ведение необходимых записей.

Основными общепринятыми формами записей являются конспект, выписки, тезисы, аннотации, резюме, план.

Конспект - это краткое письменное изложение содержания правового источника, статьи, доклада, лекции, включающее в сжатой форме основные положения и их обоснование.

Выписки - это краткие записи в форме цитат (дословное воспроизведение отрывков 12 источника, произведения, статьи, содержащих существенные положения, мысли автора), либо лаконичное, близкое к тексту изложение основного содержания.

Тезисы - это сжатое изложение ключевых идей прочитанного источника или произведения.

Аннотации, резюме - это соответственно предельно краткое обобщающее изложение содержания текста, критическая оценка прочитанного документа или произведения.

В целях структурирования содержания изучаемой работы целесообразно составлять ее план, который должен раскрывать логику построения текста, а также способствовать лучшей ориентации студента в содержании произведения.

Самостоятельная работа студентов будет эффективной и полезной в том случае, если она будет построена исходя из понимания студентами необходимости обеспечения максимально широкого охвата информационных источников.

**Рекомендации по подготовке докладов, рефератов, сообщений**

Контроль исполнения самостоятельных работ осуществляется преподавателем с участием студентов в форме обсуждения доклада, сообщения, реферата.

Тема доклада/сообщения выбирается студентом из перечней, приведенных в конце каждого раздела.

Формулировка наименования доклада согласовывается с преподавателем. Тема может быть и оригинальной, и инновационной идеей, в частности. Объем доклада должен быть таким, чтобы выступление длилось в пределах 15 минут, т.е. порядка 7-9 стр. текста шрифта 14’ через 1,5 интервала на листе А4 с полями 2 см со всех сторон.

Структура доклада: - наименование и автор, - содержание (заголовки частей), - введение (важность предлагаемой темы), - суть изложения (главные мысли и утверждения с их обоснованием), - фактический материал, факты, официальные сведения, - личное отношение докладчика к излагаемому материалу, - заключение (вывод, резюме, гипотеза, конструктивное предложение), - список использованных источников.

Конструктивным является утверждение, предложение, критика, если все они содержат действие, реализуемое в существующих условиях.

Доклад – это рационально, логично построенное повествование, имеющее целью убедить слушателей в обоснованности предлагаемых их вниманию утверждений и их следствий.

Доклад может представляться в виде презентации (PowerPoint).

Требования к презентации: - не должно быть больше семи-девяти чётких взаимосвязанных графических объектов; - не более 13 строк легко читаемого текста; - фразы должны быть лаконичными, служить сигналами докладчику в логичном изложении и слушателям в связанном восприятии; - полные скриншоты должны сопровождаться следующим слайдом с укрупнённым фрагментом, помогающим изложению; - определения можно помещать полностью или на последовательности слайдов, если строк больше 13.

Реферат представляет собой отчет студента о работе с литературой по выбранной теме.

Типовой план реферата должен включать: - тема реферата (из рекомендованных или согласованных с преподавателем); - не менее 3-х литературных источников (монографии, учебники), по каждому из которых приведена полная характеристика содержания; - материалы, выбранные из каждого источника, по теме реферата.

Примерный объем реферата – 15-20 стр., оформление как доклад.

## 6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов.

В процессе обучения используются следующие оценочные формы самостоятельной работы студентов, оценочные средства текущего контроля успеваемости и промежуточных аттестаций:

- лабораторные работы, экзамен.

### 6.1 Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине «Методы работы с большими данными»

6.1.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы.

В результате освоения дисциплины формируются следующие компетенции:

| Код компетенции | В результате освоения образовательной программы обучающийся должен обладать |
|---|---|
| ПК-2 | Способен осуществлять концептуальное, функциональное и логическое проектирование систем среднего и крупного масштаба и сложности. |

В процессе освоения образовательной программы данные компетенции, в том числе их отдельные компоненты, формируются поэтапно в ходе освоения обучающимися дисциплины в соответствии с учебным планом и календарным графиком учебного процесса.

### 6.1.2 Описание показателей и критериев оценивания компетенций,

## 7.Учебно-методическое и информационное обеспечение дисциплины

**7.1.Основная литература**

1. Изучение робототехники с помощью Python [Электронный ресурс]: Джозеф Л. Издательство "ДМК Пресс" 2019г. 250 страниц – Режим доступа https://e.lanbook.com/book/123716
2. Python и анализ данных [Электронный ресурс]: Маккинни У. Издательство "ДМК Пресс" 2020г. 540 страниц – Режим доступа: https://e.lanbook.com/book/131721

**7.2.Дополнительная литература**

1. Bhasin, H. (2019). Python Basics : A Self-Teaching Introduction. Dulles, Virginia: Mercury Learning & Information. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=edsebk&AN=1991381

## 8. Материально-техническое обеспечение дисциплины

### 8.1 Требования к оборудованию и помещению для занятий

Лабораторные работы и самостоятельная работа студентов должны проводиться в специализированной аудитории, оснащенной современной оргтехникой и персональными компьютерами с программным обеспечением в соответствии с тематикой изучаемого материала. Число рабочих мест в аудитории должно быть достаточным для обеспечения индивидуальной работы студентов. Рабочее место преподавателя должно быть оснащены современным компьютером с подключенным к нему проектором на настенный экран, или иным аналогичным по функциональному назначению оборудованием.

### 8.2 Требования к программному обеспечению

Для выполнения лабораторных работ и самостоятельной работы необходимо следующее программное обеспечение:

1. Microsoft Windows.
2. Веб-браузер, Chrome.

## 9. Методические рекомендации для самостоятельной работы студентов

## 10. Методические рекомендации для преподавателя

1. При подготовке к занятиям следует предварительно

проработать материал занятия, предусмотрев его подачу точно в отведенное для этого время занятия. Следует подготовить необходимые материалы – теоретические сведения, задачи и др. При проведении занятия следует контролировать подачу материала и решение заданий с учетом учебного времени, отведенного для занятия.

2. При проверке работ и отчетов следует учитывать не только правильность выполнения заданий, но и оптимальность выбранных методов решения, правильность выполнения всех его шагов.