|  |  |
| --- | --- |
|  | **«УТВЕРЖДАЮ»**  **Декан факультета**  **математики и компьютерных наук**  **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ /\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  подпись ФИО  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  дата |

# Рабочая программа дисциплины (модуля)

# Б1.В.ДВ.01.02 Алгоритмы обработки больших данных

(наименование и индекс дисциплины в соответствии с учебным планом)

**направление подготовки 01.04.02 "Прикладная математика и информатика"**

(код и наименование)

**направленность (профиль) «Современная теория игр»**

## Факультет математики и компьютерных наук

Кафедра прикладной математики, информационных технологий и информационной безопасности

## Рассмотрена и одобрена на заседании кафедры \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, протокол № \_\_\_\_ от «\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_ 20\_\_\_ г.

Заведующий кафедрой \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ученая степень, ученое звание, Ф.И.О., подпись)

Составитель (разработчик) программы \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ученая степень, ученое звание, Ф.И.О., подпись)

**Содержание**

|  |  |  |
| --- | --- | --- |
|  |  | стр. |
|  |  |  |
|  | Пояснительная записка |  |
|  | Цели и задачи дисциплины (модуля) |  |
|  | Объём дисциплины (модуля) по видам учебной работы |  |
|  | Содержание дисциплины (модуля) |  |
|  | Самостоятельная работа обучающихся |  |
|  | Учебно-методическое обеспечение дисциплины (модуля) |  |
|  | Методические рекомендации по дисциплине (модулю) |  |
|  | Материально-техническое обеспечение дисциплины (модуля) |  |
|  | Лист регистрации изменений |  |
|  |  |  |

**Пояснительная записка**

Рабочая программа дисциплины составлена в соответствии с требованиями ФГОС 3+ по направлению подготовки (специальности) 01.04.02 "Прикладная математика и информатика".

РП представляет собой совокупность дидактических материалов, направленных на реализацию содержательных, методических и организационных условий подготовки по направлению подготовки (специальности) 01.04.02 "Прикладная математика и информатика".

Дисциплина относится к вариативной части блока 1.

Трудоемкость дисциплины: 108 ч. / 3 з.е.;

контактная работа: 32,25 ч.

занятия лекционного типа – 10 ч.,

занятия семинарского типа (лабораторные работы) – 20 ч.,

контроль самостоятельной работы – 2 ч.,

иная контактная работа – 0,25 ч.,

контролируемая письменная работа (КПР) – 0 ч.,

СР – 75,75 ч.,

контроль – 0 ч.

Ключевые слова: большие данные, Big Data, Data Mining.

Составитель:

**1. Цели и задачи дисциплины (модуля).**

Изучение дисциплины направлено на формирование следующих компетенций:

* способностью самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять своё научное мировоззрение (ОПК-3);
* способностью углубленного анализа проблем, постановки и обоснования задач научной и проектно-технологической деятельности (ПК-3);
* способностью разрабатывать концептуальные и теоретические модели решаемых задач проектной и производственно-технологической деятельности (ПК-4);

Показателями компетенций являются:

Знания

• методы решения задач обработки и анализа больших данных, возможности высокопроизводительных вычислительных систем, технологии распределенных вычислений, методы и модели Data Mining.

Умения:

• разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных;  
• оценивать время и необходимые аппаратные ресурсы для решения задач анализа и обработки данных;   
• создавать алгоритмы анализа и обработки большого объема данных с применением моделей Data Mining.

Навыки

|  |  |
| --- | --- |
|  | • применения программных систем, предназначенных для анализа больших данных. |

**2. Объем дисциплины (модуля) по видам учебной работы.**

### Таблица 1. Объем дисциплины (модуля) общая трудоемкость: 3 з.е.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Виды учебной работы | Всего  часов | Распределение  по семестрам в часах | | | |
|  |  |  |  | III |  |
| Общая трудоемкость дисциплины | 108 |  |  | 108 |  |
| Контактная работа: | 32,25 |  |  | 32,25 |  |
| Лекции | 10 |  |  | 10 |  |
| Лабораторные работы | 20 |  |  | 20 |  |
| КСР | 2 |  |  | 2 |  |
| ИКР | 0,25 |  |  | 0,25 |  |
| Самостоятельная работа (СР) | 75,75 |  |  | 75,75 |  |
| Контроль |  |  |  |  |  |
| Вид промежуточного контроля |  |  |  | зачет |  |

**3. Содержание дисциплины (модуля).**

### Таблица 2. Распределение часов по темам и видам учебной работы

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Номер  раздела | Наименование разделов  и тем дисциплины (модуля) | Объем в часах | | | | | |
|  |  | Всего | Л | ПЗ | С | ЛР | СР  и иная работа |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| **1** | **Модуль 1** | **64** | **6** | **-** | **-** | **12** | **46** |
| 1.1 | Большие данные (Big Data): современные подходы к обработке и хранению. Проблема множественного сравнения данных. | 20 | 2 |  |  | 4 | 14 |
| 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. | 22 | 2 |  |  | 4 | 16 |
| 1.3 | Технологии KDD и Data Mining. Подготовка данных к анализу. Методика извлечения знаний. Data Mining. Мультидисциплинарный характер Data Mining. Причины распространения KDD и Data Mining. Актуальность технологий Data Mining как средств обработки больших объемов информации. | 22 | 2 |  |  | 4 | 16 |
| **2** | **Модуль 2** | **44** | **4** |  |  | **8** | **32** |
| 2.1 | Программное обеспечение в области анализа данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования. | 22 | 2 |  |  | 4 | 16 |
| 2.2 | Определение кластеризации. Постановка задачи кластеризации. Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Меры расстояний. Пример работы алгоритма k-means. Проблемы алгоритмов кластеризации. | 22 | 2 |  |  | 4 | 16 |
| **Итого** | | **108** | **10** |  |  | **20** | **78** |

**4. Самостоятельная работа обучающихся.**

### Таблица 3. Содержание самостоятельной работы обучающихся

|  |  |  |  |
| --- | --- | --- | --- |
| №  п/п | Вид самостоятельной работы | Разделы или темы  рабочей программы | Форма отчетности |
| 1 | Повторение пройденного учебного материала по конспектам лекций | Разделы 1-2 | Фронтальная беседа |
| 2 | Индивидуальные задания по теме | Разделы 1-2 | Собеседование |
| 3 | Подготовка к итоговому опросу по теме | Разделы 1-2 | Устный опрос |

**4.1. Темы курсовых работ (проектов).**

Курсовые работы или семестровые задания не предусмотрены

**4.2. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся.**

Самостоятельная работа студентов осуществляется с использованием:

1. Библиотечного фонда АГУ.

2. Компьютерного класса с доступом к сети Интернет.

**5. Учебно-методическое обеспечение дисциплины (модуля).**

### Таблица 4. Основная литература

|  |  |
| --- | --- |
| №  п/п | Наименование, библиографическое описание |
| 1 | Бутаков, Н.А. Обработка больших данных с Apache Spark : учебно-методическое пособие : [16+] / Н.А. Бутаков, М.В. Петров, Д. Насонов ; Университет ИТМО. – Санкт-Петербург : Университет ИТМО, 2019. – 52 с. : ил. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=566771> |
|  |  |

### Таблица 5. Дополнительная литература

|  |  |
| --- | --- |
| №  п/п | Наименование, библиографическое описание |
| 1 | Окулов, С.М. Алгоритмы обработки строк : учебное пособие : [12+] / С.М. Окулов. – 4-е изд., электрон. – Москва : Лаборатория знаний, 2020. – 258 с. – (Развитие интеллекта школьников). – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=445854> |
| 2 | Бродовская, Е.В. Большие данные в исследовании политических процессов : учебное пособие : [16+] / Е.В. Бродовская, А.Ю. Домбровская ; Министерство науки и высшего образования Российской Федерации, Московский педагогический государственный университет. – Москва : Московский педагогический государственный университет (МПГУ), 2018. – 88 с. : схем., табл., ил. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=563578> |

### Таблица 6. Ресурсы информационно-телекоммуникационной сети «Интернет»

|  |  |
| --- | --- |
| №  п/п | Название (адрес) ресурса |
| 1 | Data Mining - <https://www.intuit.ru/studies/courses/6/6/info> |
| 2 | Алгоритмы интеллектуальной обработки больших объемов данных - <https://www.intuit.ru/studies/courses/3498/740/info> |

**6. Методические рекомендации по дисциплине (модулю).**

Традиционно подготовка вузовской лекции строится по схеме:

- определение цели изучения материала по данной теме;

- составление плана изложения материала;

- определение основных понятий темы;

- подбор основной литературы к теме.

При подготовке лекции необходимо учитывать следующее:

1. Большое значение имеет временное планирование каждой структурной части лекции и строгое следование такому плану.

2. Необходимо максимально использовать современные технические средства обучения.

3. В случае отсутствия технических средств обучения рационально часть изучаемого материала давать через схемы, начерченные (лучше заранее) на доске. Схемы необходимо использовать для лучшего усвоения, они несут большую смысловую нагрузку.

4. Определить в процессе подготовки лекции отдельные вопросы изучаемой темы, которые будут предлагаться студентам для самостоятельного изучения.

Ведущей дидактической целью лабораторных работ является экспериментальное подтверждение и проверка существенных теоретических положений и технических сведений, изучаемых в дисциплине. Лабораторные работы должны развивать мышление студентов, самостоятельность при решении практической задачи, формировать глубоких и прочные знания.

**7. Материально-техническое обеспечение дисциплины (модуля).**

Теоретические и практические занятия должны проводиться в специализированной аудитории, оснащенной современными персональными компьютерами и программным обеспечением в соответствии с тематикой изучаемого материала. Число рабочих мест в аудитории должно быть таким, чтобы обеспечивалась индивидуальная работа студента на отдельном персональном компьютере. Аудитория также должна быть оснащенной современным видеопроектором.

Для обеспечения процесса обучения необходимо использовать помещение, рассчитанное на 12-15 слушателей и соответствующее количество лабораторных компьютеров (один компьютер на каждого учащегося).

Для выполнения лабораторных работ курса требуются компьютеры и периферийное оборудование в приведенной ниже конфигурации.

1. Операционная система MS Windows 7 или выше.

2. Пакет офисных программ Open Office или MS Office.

3. Программа просмотра PDF-документов, например, AcrobatReader.

4. Архиватор с поддержкой формата ZIP.

Рекомендуется:

* Выделенное подключение к Интернету для каждого студента, необходимое для контроля за выполнением поиска в Интернете.
* Одно многофункциональное устройство принтер/сканер/копир на класс.
* Один мультимедиа проектор на класс.

# 8. Лист регистрации изменений

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер  изменения | Номера листов | | | Основание для внесения изменения | Подпись | Расшифровка подписи | Дата | Дата  введения изменения |
|  | замененных | новых | аннулированных |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |