

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Магнитогорский государственный технический университет им. Г.И. Носова»



СВИДЕТЕЛЬСТВО:
Директор института
С. И. Лукьянов
2017 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

ТЕХНОЛОГИИ DATA MINING

Направление подготовки
09.03.01 Информатика и вычислительная техника

Профиль программы
Программное обеспечение средств вычислительной техники и
автоматизированных систем

Уровень высшего образования – бакалавриат

Программа подготовки – академический бакалавриат

Форма обучения
Очная

Институт
Кафедра
Курс
Семестр

*энергетики и автоматизированных систем
вычислительной техники и программирования*
4
8


Магнитогорск
2017 г.

Рабочая программа составлена на основе ФГОС ВО по направлению подготовки (специальности) 09.03.01 Информатика и вычислительная техника, утвержденного приказом МО и Н РФ от 12.01.2016 № 5.

Рабочая программа рассмотрена и одобрена на заседании кафедры вычислительной техники и программирования «26» сентября 2017 г., протокол № 2.


Зав. кафедрой  / О.С. Логунова/

Рабочая программа одобрена методической комиссией института энергетики и автоматизированных систем «27» сентября 2017 г., протокол № 2.

Председатель  / С.И. Лукьянов/

Рабочая программа составлена:

старшим преподавателем

 / А.Ю. Миковым

Рецензент:

начальник отдела инновационных разработок ЗАО «КонсОмСКС», канд. техн. наук


 / А.Н. Панов/

Лист актуализации рабочей программы

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2017-2018 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от 26 09 2017 г. № 2
Зав. кафедрой  О.С. Логунова

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2018 - 2019 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от 5 09 2018 г. № 1
Зав. кафедрой  О.С. Логунова

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2019 - 2020 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от 19 09 2019 г. № 5
Зав. кафедрой  О.С. Логунова

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2020 - 2021 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от 19 09 2020 г. № 5
Зав. кафедрой  О.С. Логунова

1 Цели освоения дисциплины (модуля)

Целью освоения дисциплины (модуля) «Технологии Data Mining» является формирование представления о типах задач, возникающих в области интеллектуального анализа данных (Data Mining) и методах их решения, которые помогут студентам выявлять, формализовать и успешно решать практические задачи анализа данных, возникающие в процессе их профессиональной деятельности.

Для достижения поставленной цели в курсе «Технологии Data Mining» решаются задачи:

- изучение методов статистического анализа данных;
- изучение методов интеллектуального анализа данных;
- изучение принципов организации и проведения аналитического исследования;
- решение практических задач, прогнозирование и выработка рекомендаций.

2 Место дисциплины (модуля) в структуре образовательной программы подготовки бакалавра

Дисциплина «Технологии Data Mining» входит в вариативную часть блока 1 дисциплин по выбору образовательной программы.

Для изучения дисциплины необходимы знания (умения, владения), сформированные в результате изучения курсов: программирование, базы данных, основы статистической обработки данных, обработка экспериментальных данных на ЭВМ. Обучающийся должен иметь навыки логического мышления, построения логических выводов, демонстрировать способности к использованию средств вычислительной техники к выполнению операций по обработке информации.

Знания (умения, владения), полученные при изучении данной дисциплины будут необходимы при выполнении выпускной квалификационной работы.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля) и планируемые результаты обучения

В результате освоения дисциплины (модуля) «Технологии Data Mining» обучающийся должен обладать следующими компетенциями:

Структурный элемент компетенции	Планируемые результаты обучения
ПК-3 Обладает способностью обосновывать проектные решения, осуществлять постановку и выполнять эксперименты по проверке их корректности и эффективности	
Знать	<ul style="list-style-type: none">– основные определения и понятия технологий Data Mining;– основные задачи и методы интеллектуального анализа экспериментальных данных;– основные инструменты Data Mining применяемых при проверке корректности и эффективности проектных решений.
Уметь	<ul style="list-style-type: none">– формулировать задачи Data Mining;– выбирать адекватные алгоритмы проектных решений;– оценивать корректность и эффективность проектных решений.
Владеть	<ul style="list-style-type: none">– основными методами технологий Data Mining;– практическими навыками разработки и реализации алгоритмов проектных решений с использованием технологий Data Mining;– практическими навыками разработки и реализации программных систем с использованием технологий Data Mining.

4 Структура и содержание дисциплины (модуля)

Общая трудоемкость дисциплины составляет 3 зачетных единиц 108 академических часов, в том числе:

- контактная работа – 69,95 академических часов:
 - аудиторная – 66 академических часов;
 - внеаудиторная – 3,95 академических часов
- самостоятельная работа – 2,35 академических часов;
- подготовка к экзамену – 35,7.

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в академических часах)			Самостоятельная работа (в академических часах)	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код и структурный элемент компетенции
		лекции	лаборат. занятия	практич. занятия				
Раздел 1 Обзор интеллектуальных систем	8							
1.1 Основные понятия DM и возникновение, перспективы, проблемы Data mining. Системы поддержки принятия решений (СППР). Задачи систем поддержки принятия решений. Неэффективность использования OLTP-систем для анализа данных	8	4(4И)		4	0,33	1. Подготовка к практическим занятиям 2. Самостоятельное изучение учебной и научной литературы	1. <i>Беседа - обсуждение</i> 2. <i>Проверка индивидуальных заданий</i> 3. <i>Устный опрос.</i>	ПК-3 – зув
1.2 Понятие данных. Значение понятий объект и атрибут, выборка, зависимая и независимая переменная. Типы шкал. Концепция хранилища данных. Организация ХД. Очист-	8	4(4И)		4	0,33	1. Подготовка к практическим занятиям 2. Самостоятельное изучение учебной и научной литературы	1. <i>Беседа - обсуждение</i> 2. <i>Проверка индивидуальных заданий</i> 3. <i>Устный опрос.</i>	ПК-3 – зув

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в акад. часах)			Самостоятельная работа (в акад. часах)	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код и структурный элемент компетенции
		лекции	лаборат. занятия	практич. занятия				
ка данных. Хранилища данных и анализ. Многомерная модель данных. Определение OLAP-систем. Концептуальное многомерное представление данных. Архитектура OLAP-систем: MOLAP, ROLAP, HOLAP								
Итого по разделу	8	8 (8И)		8	0,66		<i>Проверка индивидуальных заданий</i>	
Раздел 2. Методы и модели Data Mining	8							
2.1 Интеллектуальный анализ данных. Модели и методы DM. Процесс обнаружения знаний. Инструменты Data Mining. Поиск ассоциативных правил. Постановка задачи. Сиквенциальный анализ. Разновидности поиска ассоциативных правил. Методы представления результатов. Алгоритмы поиска ассоциативных правил	8	5 (5И)		5	0,338	1. Подготовка к практическим занятиям 2. Самостоятельное изучение учебной и научной литературы	1. Беседа - обсуждение 2. Проверка индивидуальных заданий 3. Устный опрос.	ПК-3 – зув

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в акад. часах)			Самостоятельная работа (в акад. часах)	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код и структурный элемент компетенции
		лекции	лаборат. занятия	практич. занятия				
2.2 Постановка задачи кластеризации. Меры близости, основанные на расстояниях, используемые в алгоритмах кластеризации. Базовые алгоритмы кластеризации. Адаптивные методы кластеризации. Классификация и регрессия. Постановка задачи. Представление результатов. Методы построения правил классификации. Методы построения математических функций. Сущность прогнозирования. Роль и значение прогнозирования. Понятия временных рядов, его компоненты и параметры. Задача визуализация данных.	8	5 (5И)		10	0,338	1. Подготовка к практическим занятиям 2. Самостоятельное изучение учебной и научной литературы	1. <i>Беседа - обсуждение</i> 2. <i>Проверка индивидуальных заданий</i> 3. <i>Устный опрос.</i>	ПК-3 – зув
2.3 Проблема распознавания образов. Классификация и характеристика основных задач распознавания образов. Формирование информативного	8	5 (5И)		5	0,338	1. Подготовка к практическим занятиям 2. Самостоятельное изучение учебной и научной литературы	1. <i>Беседа - обсуждение</i> 2. <i>Проверка индивидуальных заданий</i> 3. <i>Устный опрос.</i>	ПК-3 – зув

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в акад. часах)			Самостоятельная работа (в акад. часах)	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код и структурный элемент компетенции
		лекции	лаборат. занятия	практич. занятия				
признакового пространства. Основные понятия. Критерия информативности. Логико-эвристические методы формирования признакового пространства. Статистические методы формирования признакового пространства. Алгоритмы распознавания, основанные на вычислении оценок. Необходимые определения. Теоремы о вычислении оценок. Задачи, решаемые алгоритмами вычисления оценок.								
2.4 Мера важности объектов в сложных системах. Важность объекта в системе. Мера важности. Примеры введения меры важности объектов в сложных системах. Мера важности признаков в задаче распознавания образов. Вычислительные алгоритмы получения информационных весов и оценки	8	5 (5И)		5	0,338	1. Подготовка к практическим занятиям 2. Самостоятельное изучение учебной и научной литературы	1. <i>Беседа - обсуждение</i> 2. <i>Проверка индивидуальных заданий</i> 3. <i>Устный опрос.</i>	ПК-3 – зув

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в акад. часах)			Самостоятельная работа (в акад. часах)	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код и структурный элемент компетенции
		лекции	лаборат. занятия	практич. занятия				
их сложности. Некоторые применения мера важности признаков.								
Итого по разделу	8	25 (25И)		25	1,69		<i>Доклад с презентацией Проверка индивидуальных заданий</i>	
Итого за семестр	8	33 (33И)		33	2,35		<i>Экзамен</i>	
Итого по дисциплине	8	33 (33И)		33	2,35		<i>Экзамен</i>	

5 Образовательные и информационные технологии

1. **Традиционные образовательные технологии**, ориентированные на организацию образовательного процесса и предполагающую прямую трансляцию знаний от преподавателя к обучающемуся.

Формы учебных занятий с использованием традиционных технологий:

Информационная лекция – последовательное изложение материала в дисциплинарной логике, осуществляемое преимущественно вербальными средствами (монолог преподавателя).

Практическое занятие в форме практикума – организация учебной работы, направленная на решение комплексной учебно-познавательной задачи, требующей от студента применения как научно-теоретических знаний, так и практических навыков.

2. **Технологии проблемного обучения** – организация образовательного процесса, которая предполагает постановку проблемных вопросов, создание учебных проблемных ситуаций для стимулирования активной познавательной деятельности обучающихся.

3. **Интерактивные технологии** – организация образовательного процесса, которая предполагает активное и нелинейное взаимодействие всех участников, достижение на этой основе лично значимого для них образовательного результата.

Формы учебных занятий с использованием специализированных интерактивных технологий:

Лекция «обратной связи» – лекция–провокация (изложение материала с заранее запланированными ошибками), лекция-беседа, лекция-дискуссия, лекция-прессконференция.

4. **Информационно-коммуникационные образовательные технологии** – организация образовательного процесса, основанная на применении программных сред и технических средств работы со знаниями в различных предметных областях.

6 Учебно-методическое обеспечение самостоятельной работы обучающихся

Практические занятия

1. OLAP-системы для анализа данных.

Спроектировать хранилища данных на основании предложенной концептуальной модели базы данных:

1. Выделить измерения и факты хранилища данных.
2. Определить структуру измерений и фактов хранилища данных.
3. Проверить таблицы на предмет множественности связей между ними.
4. Проверить, все ли измерения влияют на таблицу фактов.

2. Применение методов первичного разведочного анализа данных в решении задач интеллектуального анализа данных средствами интегрированной системы Statistica.

1. Разместить в рабочей таблице пакета *Statistica* исходные эмпирические данные.
2. Для исходных эмпирических данных определить предполагаемую функцию отклика и набор факторов с обоснованием по смыслу задачи.
3. Для исходных данных выполнить построение столбчатых и круговых диаграмм, пиктографиков (3 вида), матричных графиков и контрольных карт Шухарта.
4. Подготовить описание заданий 1 и 3 в виде слайдов электронной презентации.
5. Выполнить расчет выборочных характеристик для исходных данных: обобщающие показатели, показатели вариации, показатели относительного рассеяния.
6. Построить гистограммы частот для каждого столбца, представленных исходных данных.

7. Для исходных данных выполнить:

- 1) расчет простых степенных средних, моду, медиану, показателей вариации и рассеяния;
- 2) отсев грубых погрешностей по статистике Стьюдента, при этом:
 - а) на каждом шаге итерации вычислить выборочные характеристики, обобщающие показатели, показатели вариации;
 - б) построить матрицу наблюдений после отсева;

- в) отобразить тенденцию выборочных характеристик, обобщающих показателей, показателей вариации и темп изменения каждого показателя;
- 2) проверку критериев согласия для нормального распределения для исходных данных до и после отсева, при этом:
- а) построить таблицы частот;
 - б) рассчитать значения статистик Пирсона и Колмогорова-Смирнова;
 - в) построить гистограммы частот с теоретической линией плотности нормального распределения.

3. Решение задач интеллектуального анализа данных: классификации объектов средствами интегрированной системы Statistica.

По исходным данным приведенным в выполнить:

- 1) расчет коэффициентов подобия (совстречаемости, Рао, Хаммана);
- 2) расчет расстояния Евклида и Махаланобиса;
- 3) построение матрицы парной корреляции;
- 4) кластеризацию факторов по методу корреляционных плеяд;
- 5) кластеризацию наблюдений по методу вроцлавской таксономии.

4. Применение кластеризации в решении задач интеллектуального анализа данных средствами интегрированной системы Statistica.

По исходным данным, выбранных согласно варианту выполнить:

- 1) расчет коэффициентов подобия (совстречаемости, Рао, Хаммана);
- 2) расчет расстояния Евклида и Махаланобиса;
- 3) построение матрицы парной корреляции;
- 4) кластеризацию факторов по методу корреляционных плеяд;
- 5) выполнить пошаговое отображение дендритов при кластеризации признаков и наблюдений.

5. Решение задач множественного анализа данных средствами интегрированной системы Statistica.

- 1) построение списка информативных и неинформативных факторов с оценкой мультиколлинеарности и интеркорреляции на каждом шаге;
- 2) построение множественного линейного уравнения регрессии с полным и информативным набором факторов;
- 3) сравнительный анализ построенного уравнения;
- 4) построение нелинейной регрессии согласно варианту и его оценку;
- 5) оценку предпосылок метода наименьших квадратов для линейного уравнения регрессии с полным набором факторов.

6. Задачи распознавания образов, их особенности и способы решения

1. Гистограммы изображений.
2. Выравнивание гистограмм.
3. Удаление шума и малых компонент.
4. Сглаживание изображения.
5. Фильтрация изображения.

Тестовые задания

Раздел Обзор интеллектуальных систем

- 1 Человек, который за годы обучения и практики научился чрезвычайно эффективно решать задачи, относящиеся к конкретной предметной области — ...
- 2 Эмпирическое правило, упрощающее или ограничивающее поиск решений в предметной области - ...
- 3 Внутреннее активное стремление овладеть своими собственными представлениями, понятиями, побуждениями чувств и воли, воспоминаниями, ожиданиями — ...
- 4 Проверенный практикой результат познания действительности, верное ее отражение в мышлении человека - ...

5 Переход из состояния возможности в состояние действительности, в сетевом планировании - отражение в сетевом графике выполненных работ - ...

6 Информационная база, отражающая опыт конкретных людей, человечества в целом, в решении творческих задач в выделенных сферах деятельности - база ...

7 Последовательность значений одного из анализируемых параметров многомерной базы данных - ...

8 Множественность измерений предполагает представление данных в виде «...» модели

9 Каждое измерение многомерной базы данных может быть представлено в виде «...» структуры

10 Ячейки многомерной модели данных, представленной в виде гиперкуба являются «...»

11 Среда называется, если за время между получением агентом восприятия и выработкой им решения она не изменилась

12 Среды, порождающие бесконечное число восприятий, реакций или того и другого называют

13 Среда, в которой агент формирует строго одну реакцию.

14 Среда называется, если за время между получением агентом восприятия и выработкой им решения она изменилась

15 Набор символов, принадлежащих определенному множеству

16 Раздел языка определяющий смысл этих предложений, сопоставляя символы языка с объектами реального мира, а предложения — отношения между объектами. 17 Агент состоит из знаний и , работающего с этими знаниями.

18 Таблицы в логике высказываний, позволяющие доказать общезначимость формулы называют таблицами

Раздел Методы и модели Data Mining

1. Недостатками подхода, при котором имеется хранилище данных (ХД) и витрины данных (ВД) для информации по разделам данной области, являются:

- а) увеличение объема данных, хранимых в ВД
- б) дополнительные затраты на разработку систем поддержки принятия решений с ХД и ВД
- в) увеличение нагрузки на основное ХД
- г) избыточность (данные хранятся как в ХД, так и в ВД)

2. Данные в хранилище данных делятся на следующие категории:

- а) детальные данные
- б) агрегированные данные
- в) метаданные
- г) временные данные

3. Агрегированные данные в хранилище данных подразделяются на типы:

- а) аддитивные
- б) транзитивные
- в) полуаддитивные
- г) неаддитивные

4. Метаданные описывают:

- а) объекты предметной области, информация о которых хранится в хранилище данных
- б) категории пользователей, использующих данные
- в) местоположение серверов, рабочих станций и оперативные источники данных
- г) системных администраторов
- д) размещенные на серверах и рабочих станциях программных средств и распределение данных

5. Данные, поступающие из оперативных источников данных в хранилища данных, образуют следующие информационные потоки:

- а) входной

- б) виртуальных данных
- в) метаданных
- г) обобщения

6. Семантические сети в зависимости от типа вершин различают:

- а) интенсиональные
- б) экстенсиональные
- в) рефлексивные

7. Виды условий инвариантности:

- а) гарантия
- б) выражение защищенности;
- в) выражение уверенности;
- г) качество

8. Задача классификации сводится к

- а) нахождению частых зависимостей между объектами или событиями;
- б) определению класса объекта по его характеристикам;
- в) определению по известным характеристикам объекта значения некоторого его параметра;
- г) поиску независимых групп и их характеристик во всем множестве анализируемых данных

9. Задача кластеризации заключается в:

- а) нахождении частых зависимостей между объектами или событиями;
- б) определении класса объекта по его характеристикам;
- в) определении по известным характеристикам объекта значения некоторого его параметра;
- г) поиске независимых групп и их характеристик во всем множестве анализируемых данных.

Задача регрессии сводится к:

- а) нахождению частых зависимостей между объектами или событиями;
- б) определению класса объекта по его характеристикам;
- в) определению по известным характеристикам объекта значения некоторого его параметра;
- д) поиску независимых групп и их характеристик во всем множестве анализируемых данных.

Темы рефератов

- 1 Понятие о больших данных.
- 2 Простейшие методы обработки.
- 3 Многомерные статистические методы в экономике, управлении и финансах.
- 4 Реализация статистических методов в пакетах прикладных программ.
- 5 Математические основы многомерных статистических методов.
- 6 Методы множественного корреляционно-регрессионного анализа.
- 7 Постановка задач классификации.
- 8 Кластерный анализ.
- 9 Использование кластерного анализа.
- 10 Постановка задач снижения размерности.
- 11 Использование компонентного анализа.
- 12 Факторный анализ.
- 13 Использование факторного анализа.
- 14 Введение в теорию качественных признаков и нечисловой информации.
- 15 Методы оцифровки.
- 16 Методы Data Mining.
- 17 Новые технологии обработки и хранения больших данных.
- 19 Технологии поиска данных.
- 20 Интеграция данных из различных источников.

7 Оценочные средства для проведения промежуточной аттестации

а) Планируемые результаты обучения и оценочные средства для проведения промежуточной аттестации:

Структурный элемент компетенции	Планируемые результаты обучения	Оценочные средства
ПК-3 Способностью обосновывать проектные решения, осуществлять постановку и выполнять эксперименты по проверке их корректности и эффективности		
Знать	<ul style="list-style-type: none"> – основные определения и понятия технологий Data Mining; – основные задачи и методы интеллектуального анализа экспериментальных данных; – основные инструменты Data Mining применяемых при проверке корректности и эффективности проектных решений. 	<p><i>Перечень теоретических вопросов</i></p> <ol style="list-style-type: none"> 1. Задача анализа данных. Интеллектуальный анализ данных. Отличия от других видов анализа данных. 2. Постановка и порядок решение задачи интеллектуального анализа данных на примере информационной модели торговой компании. 3. Закономерности в данных: неочевидные, объективные, полезные. Области применения интеллектуального анализа данных. 4. Подготовка данных для анализа. Обработка пропущенных и недостающих данных. Анализ исключений. 5. Основные концепции баз данных. Модель данных. СУБД. Использование баз данных в интеллектуальном анализе данных. 6. Постановка и порядок решения задачи интеллектуального анализа данных на примере хранилища слабоструктурированных текстовых документов. 7. Введение в многомерные базы данных. Особенности базы данных для оперативной аналитической обработки данных. Порядок и особенности построения хранилищ данных. 8. Виды и особенности шкал измерений данных. 9. Введение в OLAP. Типовая модель данных для OLAP. Особенности приложений для оперативной аналитической обработки данных. 10. Постановка и порядок решения задачи интеллектуального анализа данных на примере документной базы данных. 11. Особенности использования данных при интеллектуальном анализе данных. Данные, информация и знания. 12. Основные задачи интеллектуального анализа данных. 13. Стадии интеллектуального анализа данных. Начальные этапы. Вычисле-

Структурный элемент компетенции	Планируемые результаты обучения	Оценочные средства
		<p>ние и анализ показателей описательной статистики.</p> <p>14. Характеристики временных рядов, используемые в ходе интеллектуального анализа данных.</p> <p>15. Построение и использование моделей для интеллектуального анализа данных.</p> <p>16. Решение задач классификации в ходе интеллектуального анализа данных: цель и основные методы.</p> <p>17. Преимущественные особенности методов анализа данных: статистических, машинного обучения, оперативного анализа, интеллектуального анализа.</p> <p>18. Решение задач кластеризация в ходе интеллектуального анализа данных: цель и основные методы.</p> <p>19. Сущность и методы решения задачи прогнозирования при интеллектуальном анализе данных.</p> <p>20. Комплексный подход к внедрению Data Mining.</p>
Уметь	<ul style="list-style-type: none"> – формулировать задачи Data Mining; – выбирать адекватные алгоритмы проектных решений; – оценивать корректность и эффективность проектных решений. 	<p><i>Практические задания</i></p> <ol style="list-style-type: none"> 1. Изучить возможности и освоить основные принципы работы в интегрированной статистической системе Statistica; 2. Получить практические навыки создания, анализа, визуализации данных в Statistica; 3. Изучить средства Statistica для проведения первичного разведочного анализа данных (методы визуализации и аналитические методы) на примере решения конкретной задачи 4. Изучить алгоритмы и методы кластерного и дискриминантного анализов данных на примере решения конкретной задачи; 5. Исследовать эффективность использования различных алгоритмов и методов кластерного анализа данных для решения прикладной задачи; 6. Ознакомиться и получить практические навыки работы с модулями интегрированной статистической системы Statistica, реализующими решение задачи классификации объектов.

Структурный элемент компетенции	Планируемые результаты обучения	Оценочные средства
		<p>7. Изучить алгоритмы и методы кластерного и дискриминантного анализов данных на примере решения конкретной задачи;</p> <p>8. Исследовать эффективность использования различных алгоритмов и методов кластерного анализа данных для решения прикладной задачи;</p> <p>9. Ознакомиться и получить практические навыки работы с модулями интегрированной статистической системы Statistica, реализующими решение задачи классификации объектов.</p> <p>10. Изучить методы и алгоритмы прогнозирования временных рядов на примере решения конкретной задачи.</p>
Владеть	<ul style="list-style-type: none"> – основными методами технологий Data Mining; – практическими навыками разработки и реализации алгоритмов проектных решений с использованием технологий Data Mining; – практическими навыками разработки и реализации программных систем с использованием технологий Data Mining. 	<p><i>Задания на решение задач из профессиональной области, комплексные задания</i></p> <p>Реализовать, используя механизм общей памяти, передачу данных между двумя потоками. Первый поток проводит вычислительную работу, оперируя данными, считываемых из файла. Результат вычислений для каждого цикла вычислений появляется асинхронно. Время появления результата на каждом цикле вычислений зависит от множества различных факторов. При появлении результата вычислений первый поток должен передавать его второму потоку для дальнейших вычислений. Исходный файл – содержит ~10000 строк чисел типа float. Первый поток считывает их последовательно по ~1000 и производит над ними вычислительные операции (например нахождение среднеквадратичного). Результат каждого вычислительного цикла передается второму потоку. Второй поток также производит над ним мат. операции и конечный результат пишет в файл.</p> <p>Требования реализации:</p> <ul style="list-style-type: none"> - использование системных объектов синхронизации; - все записи считывания и записи в файл дублировать на консоль (форму); - моменты переключения между потоками сообщать на консоль (форму); - описание программы.

б) Порядок проведения промежуточной аттестации, показатели и критерии оценивания:

Промежуточная аттестация по дисциплине «Технологии Data Mining» включает теоретические вопросы, позволяющие оценить уровень усвоения обучающимися знаний, и практические задания, выявляющие степень сформированности умений и владений, проводится в форме экзамена.

Экзамен по данной дисциплине проводится в устной форме по экзаменационным билетам, каждый из которых включает 2 теоретических вопроса и одно практическое задание

Показатели и критерии оценивания зачета с оценкой:

– на оценку «**отлично**» (5 баллов) – обучающийся демонстрирует высокий уровень сформированности компетенций, всестороннее, систематическое и глубокое знание учебного материала, свободно выполняет практические задания, свободно оперирует знаниями, умениями, применяет их в ситуациях повышенной сложности.

– на оценку «**хорошо**» (4 балла) – обучающийся демонстрирует средний уровень сформированности компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.

– на оценку «**удовлетворительно**» (3 балла) – обучающийся демонстрирует пороговый уровень сформированности компетенций: в ходе контрольных мероприятий допускаются ошибки, проявляется отсутствие отдельных знаний, умений, навыков, обучающийся испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.

– на оценку «**неудовлетворительно**» (2 балла) – обучающийся демонстрирует знания не более 20% теоретического материала, допускает существенные ошибки, не может показать интеллектуальные навыки решения простых задач.

– на оценку «**неудовлетворительно**» (1 балл) – обучающийся не может показать знания на уровне воспроизведения и объяснения информации, не может показать интеллектуальные навыки решения простых задач.

8 Учебно-методическое и информационное обеспечение дисциплины (модуля)

а) Основная литература:

1. Шитиков В.К. Классификация, регрессия и другие алгоритмы Data Mining с использованием R [Электронный ресурс]./ В.К. Шитиков, С.Э. Мастицкий - Тольятти Лондон, -2017, 351 с. Режим доступа: <https://ranalytics.github.io/data-mining> .

б) Дополнительная литература:

1. Радченко И.А. Технологии и инфраструктура Big Data: Учебное пособие. [Электронный ресурс]. / И.А. Радченко, И.Н. Николаев – СПб.: Университет ИТМО, 2018. – 55 с. Режим доступа

http://books.ifmo.ru/book/2138/tehnologii_i_infrastructura_Big_Data:_uchebnoe_posobie.htmhttp://books.ifmo.ru/book/2138/tehnologii_i_infrastructura_Big_Data:_uchebnoe_posobie.htm

в) Программное обеспечение и Интернет-ресурсы:

Программное обеспечение: лицензионное программное обеспечение: операционная система; офисные программы; математические пакет, статистические пакеты, установленные на каждом персональном компьютере вычислительного центра ФГБОУ ВПО «МГТУ».

Перечень лицензионного программного обеспечения по ссылке:

<http://sps.vuz.magtu.ru/Shared%20Documents/Forms/AllItems.aspx?RootFolder=%2FShared%20Documents%2F%D0%9F%D0%BE%D0%B4%D0%B3%D0%BE%D1%82%D0%BE%D0%B2%D0%BA%D0%B0%20%D0%BA%20%D0%B0%D0%BA%D0%BA%D1%80%D0%B5%D0%B4%D0%B8%D1%82%D0%B0%D1%86%D0%B8%D0%B8%202020%2F%D0%A1%D0%B0%D0%BC%D0%BE%D0%BE%D0%B1%D1%81%D0%BB%D0%B5%D0%B4%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5%202019%D0%B3%2F%D0%9B%D0%B8%D1%86%D0%B5%D0%BD%D0%B7%D0%B8%D0%BE%D0%BD%D0%BD%D0%BE%D0%B5%20%D0%9F%D0%9E&InitialTabId=Ribbon.Document&VisibilityContext=WSSTabPersistence>

Официальные сайты промышленных предприятий и организаций: <http://www.mmk.ru>, <http://www.creditural.ru>, <http://www.magtu.ru>, <http://www.gks.ru> и т.п.; разработчиков

программных продуктов: <http://www.statsoft.ru>, <http://www.microsoft.com>, <http://www.ptc.com> и т.п.; сайты лабораторий компьютерной графики <http://graphics.cs.msu.ru>, <http://cgm.graphicon.ru>.

9 Материально-техническое обеспечение дисциплины (модуля)

Материально-техническое обеспечение дисциплины включает:

Тип и название аудитории	Оснащение аудитории
Лекционная аудитория	Мультимедийные средства хранения, передачи и представления информации
Компьютерный класс	Персональные компьютеры с пакетом Office, выходом в Интернет и с доступом в электронную информационно-образовательную среду университета
Аудитории для самостоятельной работы: компьютерные классы; читальные залы библиотеки	Все классы УИТ и АСУ с персональными компьютерами, выходом в Интернет и с доступом в электронную информационно-образовательную среду университета
Аудиторий для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации	Ауд. 282 и классы УИТ и АСУ
Помещения для самостоятельной работы обучающихся, оснащенных компьютерной техникой с возможностью подключения к сети «Интернет» и наличием доступа в электронную информационно-образовательную среду организации	Классы УИТ и АСУ
Помещения для хранения и профилактического обслуживания учебного оборудования	Центр информационных технологий – ауд. 379