

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**

Методы и инструменты анализа больших данных

Код модуля
1156039(1)

Модуль
Математические методы информационной
безопасности

Екатеринбург

Оценочные материалы составлены автором(ами):

| № п/п | Фамилия, имя, отчество | Ученая степень, ученое звание | Должность | Подразделение |
|--------------|-------------------------------|---|-----------------------|----------------------|
| 1 | Бородин Андрей Михайлович | кандидат технических наук, без ученого звания | Доцент | |
| 2 | Мирвода Сергей Геннадьевич | без ученой степени, без ученого звания | Старший преподаватель | |
| 3 | Пономарева Ольга Алексеевна | кандидат технических наук, без ученого звания | Доцент | |

Согласовано:

Управление образовательных программ

Т.Г. Комарова

Авторы:

- Бородин Андрей Михайлович, Доцент,
- Мирвода Сергей Геннадьевич, Старший преподаватель,
- Пономарева Ольга Алексеевна, Доцент,

1. СТРУКТУРА И ОБЪЕМ ДИСЦИПЛИНЫ **Методы и инструменты анализа больших данных**

| | | | |
|----|--------------------------------------|--------------------------------|---|
| 1. | Объем дисциплины в зачетных единицах | 3 | |
| 2. | Виды аудиторных занятий | Лекции Лабораторные занятия | |
| 3. | Промежуточная аттестация | Зачет | |
| 4. | Текущая аттестация | Контрольная работа | 1 |
| | | Домашняя работа | 1 |

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ МОДУЛЯ **Методы и инструменты анализа больших данных**

Индикатор – это признак / сигнал/ маркер, который показывает, на каком уровне обучающийся должен освоить результаты обучения и их предъявление должно подтвердить факт освоения предметного содержания данной дисциплины, указанного в табл. 1.3 РПМ-РПД.

Таблица 1

| Код и наименование компетенции | Планируемые результаты обучения (индикаторы) | Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине |
|--|--|---|
| 1 | 2 | 3 |
| УК-1 -Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий | З-1 - Определять принципы критического мышления, методов анализа и оценки информации П-1 - Выявлять и анализировать проблемную ситуацию, выделяя ее структурные составляющие и связи между ними У-1 - Осмысливать явления окружающего мира во взаимосвязи, целостности и развитии, выстраивать логические связи между элементами системы | Домашняя работа Зачет Контрольная работа Лабораторные занятия Лекции |
| ОПК-4 -Способен осуществлять сбор, | З-2 - Понимать методы анализа и обоснования выбора решений | Домашняя работа Зачет |

| | | |
|--|--|---|
| <p>обработку и анализ научно-технической информации по теме исследования, разрабатывать планы и программы проведения научных исследований и технических разработок</p> | <p>по обеспечению требуемого уровня безопасности информационных систем 3-4 - Объяснять правила, способы и методы организации, выполнения и представления результатов научного исследования 3-5 - Понимать основные категории и понятия информационно аналитической работы, принципы и методы ее ведения 3-6 - Понимать методы выработки и принятия информационного решения 3-7 - Использовать технологии поиска, изучения, обобщения и систематизации научной информации П-1 - Пользоваться навыками структурирования информации по теме исследования П-4 - Осуществлять поиск информации в глобальной информационной сети Интернет У-3 - Использовать программные и аппаратные средства персонального компьютера для поиска и обработки информации</p> | <p>Контрольная работа Лабораторные занятия Лекции</p> |
|--|--|---|

3. ПРОЦЕДУРЫ КОНТРОЛЯ И ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ В РАМКАХ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ МОДУЛЯ В БАЛЬНО-РЕЙТИНГОВОЙ СИСТЕМЕ (ТЕХНОЛОГИЧЕСКАЯ КАРТА БРС)

3.1. Процедуры текущей и промежуточной аттестации по дисциплине

| | | |
|---|--|-------------------------------------|
| 1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.50 | | |
| Текущая аттестация на лекциях | Сроки – семестр, учебная неделя | Максимальная оценка в баллах |
| <i>контрольная работа</i> | 3,4 | 100 |
| Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.50 | | |
| Промежуточная аттестация по лекциям – зачет | | |

| | | |
|---|---------------------------------|------------------------------|
| Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.50 | | |
| 2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – не предусмотрено | | |
| Текущая аттестация на практических/семинарских занятиях | Сроки – семестр, учебная неделя | Максимальная оценка в баллах |
| | | |
| Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям– не предусмотрено | | |
| Промежуточная аттестация по практическим/семинарским занятиям–нет | | |
| Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям– не предусмотрено | | |
| 3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий –0.50 | | |
| Текущая аттестация на лабораторных занятиях | Сроки – семестр, учебная неделя | Максимальная оценка в баллах |
| <i>домашняя работа</i> | 3,14 | 100 |
| Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям -1.00 | | |
| Промежуточная аттестация по лабораторным занятиям –нет | | |
| Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – 0.00 | | |
| 4. Онлайн-занятия: коэффициент значимости совокупных результатов онлайн-занятий –не предусмотрено | | |
| Текущая аттестация на онлайн-занятиях | Сроки – семестр, учебная неделя | Максимальная оценка в баллах |
| | | |
| Весовой коэффициент значимости результатов текущей аттестации по онлайн-занятиям -не предусмотрено | | |
| Промежуточная аттестация по онлайн-занятиям –нет | | |
| Весовой коэффициент значимости результатов промежуточной аттестации по онлайн-занятиям – не предусмотрено | | |

3.2. Процедуры текущей и промежуточной аттестации курсовой работы/проекта

| | | |
|---|---------------------------------|------------------------------|
| Текущая аттестация выполнения курсовой работы/проекта | Сроки – семестр, учебная неделя | Максимальная оценка в баллах |
| | | |
| Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– не предусмотрено | | |
| Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – не предусмотрено | | |

4. КРИТЕРИИ И УРОВНИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ МОДУЛЯ

4.1. В рамках БРС применяются утвержденные на кафедре/институте критерии (признаки) оценивания достижений студентов по дисциплине модуля (табл. 4) в рамках контрольно-оценочных мероприятий на соответствие указанным в табл.1 результатам обучения (индикаторам).

Таблица 4

Критерии оценивания учебных достижений обучающихся

| Результаты обучения | Критерии оценивания учебных достижений, обучающихся на соответствие результатам обучения/индикаторам |
|----------------------------|--|
| Знания | Студент демонстрирует знания и понимание в области изучения на уровне указанных индикаторов и необходимые для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью. |
| Умения | Студент может применять свои знания и понимание в контекстах, представленных в оценочных заданиях, демонстрирует освоение умений на уровне указанных индикаторов и необходимых для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью. |
| Опыт /владение | Студент демонстрирует опыт в области изучения на уровне указанных индикаторов. |
| Другие результаты | Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов. Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения. Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения. |

4.2 Для оценивания уровня выполнения критериев (уровня достижений обучающихся при проведении контрольно-оценочных мероприятий по дисциплине модуля) используется универсальная шкала (табл. 5).

Таблица 5

Шкала оценивания достижения результатов обучения (индикаторов) по уровням

| Характеристика уровней достижения результатов обучения (индикаторов) | | | | |
|---|---|---|---------|---|
| № п/п | Содержание уровня выполнения критерия оценивания результатов обучения (выполненное оценочное задание) | Шкала оценивания | | |
| | | Традиционная характеристика уровня | | Качественная характеристика уровня |
| 1. | Результаты обучения (индикаторы) достигнуты в полном объеме, замечаний нет | Отлично (80-100 баллов) | Зачтено | Высокий (В) |
| 2. | Результаты обучения (индикаторы) в целом достигнуты, имеются замечания, которые не требуют обязательного устранения | Хорошо (60-79 баллов) | | Средний (С) |

| | | | | |
|----|--|--|------------|-------------------|
| 3. | Результаты обучения (индикаторы) достигнуты не в полной мере, есть замечания | Удовлетворительно (40-59 баллов) | | Пороговый (П) |
| 4. | Освоение результатов обучения не соответствует индикаторам, имеются существенные ошибки и замечания, требуется доработка | Неудовлетворительно (менее 40 баллов) | Не зачтено | Недостаточный (Н) |
| 5. | Результат обучения не достигнут, задание не выполнено | Недостаточно свидетельств для оценивания | | Нет результата |

5. СОДЕРЖАНИЕ КОНТРОЛЬНО-ОЦЕНОЧНЫХ МЕРОПРИЯТИЙ ПО ДИСЦИПЛИНЕ МОДУЛЯ

5.1. Описание аудиторных контрольно-оценочных мероприятий по дисциплине модуля

5.1.1. Лекции

Самостоятельное изучение теоретического материала по темам/разделам лекций в соответствии с содержанием дисциплины (п. 1.2. РПД)

5.1.2. Лабораторные занятия

Примерный перечень тем

1. Установка сервера Hadoop или подключение к облачному сервису. Загрузка тестового набора данных. Выполнение задачи тестовой задачи MapReduce. Написание собственной задачи на языке Java.

2. Установка сервера ClickHouse, Оценка и подбор табличного движка под условия задачи. Создание таблицы и загрузка в неё тестового набора данных. Написание выборок на языке SQL.

3. Установка сервера redis, подбор типов данных под условия задачи. Загрузка тестового набора данных. Написание выборок на языке C#.

4. Установка или использование облачных сервисов Apache Spark. Работа с RDD данными. Трансформации и вычисления.

5. Основы языка R: Манипулирование объектами данных; использование встроенных функций; построение простейших графиков; простейшие статистические концепции.

6. Расширения языка R: работа с текстом; использование регулярных выражений для извлечения данных; переписывание существующего кода; ассиметричные распределения; основы исследовательского анализа данных.

7. Библиотеки машинного обучения и получения данных из ХД; библиотеки RHIPE и RHadoop; генерирование рекомендаций про помощи данных из выбранного ХД и методов машинного обучения.

LMS-платформа – не предусмотрена

5.2. Описание внеаудиторных контрольно-оценочных мероприятий и средств текущего контроля по дисциплине модуля

Разноуровневое (дифференцированное) обучение.

Базовый

5.2.1. Контрольная работа

Примерный перечень тем

1. Анализ данных: ассоциации – поиск закономерностей внутри набора одновременно наступивших событий.

2. Анализ данных: последовательность – поиск причинно-следственной связи между последовательно случившимися событиями.

3. Анализ данных: определение отклонений – нахождение данных существенно отличающихся от большинства, поиск характерных шаблонов данных.

4. Методы исследовательского анализа: графические инструменты нахождения зависимостей, выбросов.

Примерные задания

1. Основные статистические методы анализа данных:

a. Классификация – задача обнаружения признаков характеризующих группу объектов на основании сохранённой в ХД информации.

b. Кластеризация – разбиение объектов на группы по заранее не известным признакам

c. Прогнозирование – нахождение пропущенных или последующих значений показателей

d. Оценивание

Автоматическое разбиение элементов некоторого множества (объекты, данные, вектора характеристик) на группы по принципу схожести:

a. кластеризация

б. классификация

в. интерполяция

г. модуляция

Какие из алгоритмов относятся к алгоритмам кластеризации:

a. Метод ближайшего соседа

б. Минимальное покрывающее дерево

в. k-Means алгоритм

г. Генетические алгоритмы

Расположите в правильном порядке уровни современной информационно-аналитической системы а.извлечение, преобразование и загрузка данных

б. сбор и первичная обработка данных

с. складирование данных

д. анализ данных

е. представление данных в витринах данных

ф. Web-портал:

a. a,b,c,d,e,f

б. b,c,a,f,d,e

в. a,c,b,f,e,d

г. b,a,c,e,d,f

Процесс превращения данных в знания, а знаний в действия бизнеса для получения выгоды. Является деятельностью конечного пользователя, которую облегчают различные аналитические и групповые инструменты и приложения, а также инфраструктура хранилища данных:

- a. Business Intelligence
- б. Свободный поиск
- в. Анализ исключений
- г. Классификация

Выберите верные утверждения:

- a. Данные в традиционной базе данных структурированы.
 - б. Данные в базе больших данных структурированы
 - в. Данные в традиционной базе данных полу-структурированы или неструктурированные
 - г. Данные в базе больших данных полу-структурированы или неструктурированные
- Этапы обучения в правильном порядке
- a. Сбор данных;
 - б. Подготовка данных (фильтрация, дополнение, кодирование);
 - в. Постановка задачи анализа;
 - г. Подбор параметров модели и алгоритма обучения;
 - д. Выбор модели (алгоритма анализа данных);
 - е. Обучение модели (автоматический поиск остальных параметров модели);
 - ж. Анализ качества обучения, если неудовлетворительный переход на п. 5 или п. 4;
 - з. Анализ выявленных закономерностей, если неудовлетворительный переход на п. 1, 4 или 5.

- a. c,a,b,e,d,f,g,h
- б. a,b,c,d,e,f,g,h
- в. b,c,a,e,f,h,g,d

г. нет правильного ответа

Классификация методов Data Mining по задачам

a. задачи классификации и кластеризации

б. задачи прогнозирования

в. 1,2

г. нет верного

LMS-платформа – не предусмотрена

5.2.2. Домашняя работа

Примерный перечень тем

1. Аналитическая обработка больших данных предметной области

Примерные задания

.1. Исследование особенностей документных баз данных

2. Исследование особенностей графовых баз данных

3. Исследование особенностей баз данных семейства колонок

4. Исследование особенностей построения систем больших данных на основе обработки потоков данных
5. Использование языка Python для построения аналитических запросов к ХД BigData.
6. Изучение библиотек Python для анализа данных.
7. Исследование методов расширения функциональности классических РСУБД для построения на их основе систем анализа больших данных.
8. Оформить отчет по домашней работе
LMS-платформа – не предусмотрена

5.3. Описание контрольно-оценочных мероприятий промежуточного контроля по дисциплине модуля

5.3.1. Зачет

Список примерных вопросов

1. Укажите принципиальные характеристики “Новых” данных.
 2. Какие бывают типы данных.
 3. Назовите типы ограничений (constraints) СУБД.
 4. Что такое структурированные данные.
 5. Что такое не структурированные данные.
 6. Что такое полу-структурированные данные.
 7. Приведите определение строк, столбцов, таблиц реляционной базы данных.
 8. Опишите характеристики ACID баз данных.
 9. Приведите ключевую концепцию реляционных баз данных.
 10. Приведите различные шаги по работе с неструктурированными данными.
- LMS-платформа – не предусмотрена

5.4 Содержание контрольно-оценочных мероприятий по направлениям воспитательной деятельности

Направления воспитательной деятельности сопрягаются со всеми результатами обучения компетенций по образовательной программе, их освоение обеспечивается содержанием всех дисциплин модулей.