

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
"Казанский (Приволжский) федеральный университет"
филиал в городе Каире Арабской Республики Египет



УТВЕРЖДАЮ

Ректор

Д.Р. Сафин

2024 г.
МП



Программа дисциплины (модуля)

Исследование данных

Направление подготовки: 09.03.04 – Программная инженерия

Профиль подготовки: Искусственный интеллект и современная разработка программного обеспечения

Квалификация выпускника: бакалавр

Форма обучения: очная

Язык обучения: английский

Год начала обучения по образовательной программе: 2025

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-8	Владение навыками использования различных технологий разработки программного обеспечения и систем искусственного интеллекта

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

- 1) Модели данных
- 2) Работа с данными через SQL
- 3) Работа с данными через Python
- 4) Работа с большими данными , используя паттерн MapReduce

Должен уметь:

- 1) Подготавливать / очищать / нормализовать данные перед исследованием
- 2) Делать анализ требований на основании текущей задачи
- 3) Применять методы и алгоритмы для решения поставленной задачи

Должен владеть:

- 1) Навыками SQL и написания обработки данных на Python
- 2) Создавать MVP проекта на основании алгоритмов DataScience

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в Блок «Дисциплины, модули» Б1.В.ДВ.04.08 основной профессиональной образовательной программы 09.03.04 "Программная инженерия (Современная разработка программного обеспечения)" и относится к части, формируемой участниками образовательных отношений.

Осваивается на 3 курсе в 5 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины (модуля) составляет 5 зачетных(ые) единиц(ы) на 180 часа(ов).

Контактная работа - 72 часа(ов), в том числе лекции - 0 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 72 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 72 часа(ов).

Контроль (зачёт / экзамен) - 36 часа(ов).

Форма промежуточного контроля дисциплины (модуля): экзамен в 5 семестре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

N	Разделы дисциплины (модуля)	С е м е с т р	Виды и часы контактной работы, их трудоемкость (в часах)			Самост оятельн ая работа
			Лек ции	Практ ическ ие заняти я	Лабора торн ые работ ы	
1.	Тема 1: Введение в Data Mining	5	0	0	9	9
2.	Тема 2: Подготовка и нормализация данных	5	0	0	9	9
3.	Тема 3: Метрики схожести объектов	5	0	0	9	9
4.	Тема 4: Ранжирование и анализ ссылок	5	0	0	9	9
5.	Тема 5: Алгоритмы работы с рекламой в интернете	5	0	0	9	9
6.	Тема 6: Задача частотности повторений для рекомендательных систем	5	0	0	9	9
7.	Тема 7: Алгоритмы работы с потоковыми данными	5	0	0	9	9
8.	Тема 8: Кластеризация	5	0	0	9	9
	Итого		0	0	72	72

4.2 Содержание дисциплины (модуля)

Тема 1. Введение в Data Mining

- данные
- эволюция данных
- принцип Бонферрони
- типы данных для матрицы данных
- общие задачи и паттерны алгоритмов
- структурированные и неструктурированные данные

Разбор практических примеров для работы с разными типами данных

Тема 2. Подготовка и нормализация данных

- очистка данных
- нормализация данных
- приведение данных к согласованному виду
- качество данных

Разбор практических примеров для работы с нормализацией данных

Тема 3. Метрики схожести объектов

- дистанция между объектами в неевклидовом пространстве
- норма расстояния
- расстояние Минковского / Махаланобиса
- матрица схожести
- коэффициент Жаккара

Разбор практических примеров для работы с вычислением расстояния между объектами

Тема 4. Ранжирование и анализ ссылок

- алгоритм PageRank
- алгоритм TrustRank
- алгоритм HITS
- анализ SPAM фермы

Разбор практических примеров для работы с вычислением ранжирования Web страниц

Тема 5. Алгоритмы работы с рекламой в интернете

- жадные алгоритмы
- алгоритм баланса
- ключевые слова и проведение аукционов

Разбор практических примеров для работы с использованием жадных алгоритмов

Тема 6. Задача частотности повторений для рекомендательных систем

- ассоциативные правила
- метрики анализа ассоциаций
- принцип априори
- семейство хеш алгоритмов для работы с частотными данными

Разбор практических примеров для работы с использованием алгоритмов априори

Тема 7. Алгоритмы работы с потоковыми данными

- жадные алгоритмы в потоковых данных
- вероятностные структуры семейства Блум фильтров
- алгоритм LogLog / HyperLogLog

Разбор практических примеров для работы с использованием анализа потоковых данных

Тема 8. Кластеризация

- стратегии кластеризации
- алгоритмы иерархической кластеризации
- метрики кластеров
- другие виды алгоритмов работы с кластерами данных

Разбор практических примеров для работы с построением кластеризации на данных

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-996ин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

6. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;

– содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

– в электронном виде – через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;

– в печатном виде – в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе «Электронный университет». При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину (модуль).

Перечень литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины (модуля). Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

1. Charu C. Aggarwal. Data Mining. – URL: <https://link.springer.com/content/pdf/10.1007/978-3-319-14142-8.pdf>
2. Charu C. Aggarwal. Outlier Analysis. – URL: <https://link.springer.com/content/pdf/10.1007/978-3-319-47578-3.pdf>
3. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman. Mining of Massive Datasets. – URL: <http://infolab.stanford.edu/~ullman/mmds/book.pdf>
4. Jiawei Han, Micheline Kamber, Jian Pei. Data Mining Concepts and Techniques. – URL: <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>
5. Habrahabr - <https://habrahabr.ru>
6. <https://datalaboratory.one>

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Вид работ	Методические рекомендации
лабораторные работы	Для успешной подготовки к лабораторной работе обучающийся должен ознакомиться с материалами предоставленных лекций Список лабораторных работ приведен ниже: - нормализация и подготовка данных - использование алгоритмов построения ранжирования объектов

Вид работ	Методические рекомендации
	<ul style="list-style-type: none"> - использование жадных алгоритмов для обработки потоковых данных - использование алгоритмов кластеризации
самостоятельная работа	<p>Важнейшим этапом практического занятия является самостоятельная работа обучающихся, которая складывается из нескольких разделов:</p> <ol style="list-style-type: none"> 1. Теоретическая самоподготовка обучающихся по некоторым учебным темам, входящим в примерный тематический учебный план 2. Знакомство с дополнительной учебной литературой и другими учебными методическими материалами, закрепляющими некоторые практические навыки обучающихся (учебными материалами в интернете, онлайн-курсами).
экзамен	<p>Для успешной подготовки к экзамену обучающийся должен самостоятельно выполнить все лабораторные работы и программы по основным темам курса.</p>

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

- помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду КФУ;
- учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья);
- компьютер и принтер для распечатки раздаточных материалов;
- проектор.

12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;

- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;

- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;

- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной (модулем), за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;

- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;

- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;

- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:

- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;

- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;

- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 09.03.04 "Программная инженерия (Современная разработка программного обеспечения).

Приложение №1
к рабочей программе дисциплины (модуля)
Б1.В.ДВ.04.08. Исследование данных

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«Казанский (Приволжский) федеральный университет»
Филиал в городе Каире Арабской Республики Египет

**Фонд оценочных средств по дисциплине (модулю)
Б1.В.ДВ.04.08. Исследование данных**

Направление подготовки: 09.03.04 – Программная инженерия
Профиль подготовки: Искусственный интеллект и современная разработка программного обеспечения
Квалификация выпускника: бакалавр
Форма обучения: очная
Язык обучения: английский
Год начала обучения по образовательной программе: 2025

СОДЕРЖАНИЕ

1. СООТВЕТСТВИЕ КОМПЕТЕНЦИЙ ПЛАНИРУЕМЫМ РЕЗУЛЬТАТАМ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

2. КРИТЕРИИ ОЦЕНИВАНИЯ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ

3. РАСПРЕДЕЛЕНИЕ ОЦЕНОК ЗА ФОРМЫ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНУЮ АТТЕСТАЦИЮ

4. ОЦЕНОЧНЫЕ СРЕДСТВА, ПОРЯДОК ИХ ПРИМЕНЕНИЯ И КРИТЕРИИ ОЦЕНИВАНИЯ

4.1. ОЦЕНОЧНЫЕ СРЕДСТВА ТЕКУЩЕГО КОНТРОЛЯ

4.1.1. Лабораторные работы

4.1.1.1. Порядок проведения и процедура оценивания

4.1.1.2. Критерии оценивания

4.1.1.3. Содержание оценочного средства

4.1.2. Контрольная работа

4.1.2.1. Порядок проведения и процедура оценивания

4.1.2.2. Критерии оценивания

4.1.2.3. Содержание оценочного средства

4.2. Оценочные средства промежуточной аттестации

4.2.1. Экзамен. Устный ответ

4.2.1.1. Порядок проведения

4.2.1.2. Критерии оценивания

4.2.1.3. Оценочные средства

1. Соответствие компетенций планируемым результатам обучения по дисциплине (модулю)

Код и наименование компетенции	Индикаторы достижений компетенций	Оценочные средства текущего контроля и промежуточной аттестации
<p>ПК-8 Владение навыками использования различных технологий разработки программного обеспечения и систем искусственного интеллекта</p>	<p>Знать основные понятия и формулировки в области применения алгоритмов исследования данных, основные методы и паттерны применяемые для задач, связанных с получением информации из данных.</p> <p>Уметь осуществлять расчет на основании выбора алгоритма исследований данных, осуществлять расчет подготовленных и нормализованных данных</p> <p>Владеть навыками применения алгоритмов исследования данных, владеет навыками расчета подготовленных и нормализованных данных</p>	<p>Текущий контроль: Лабораторные работы Контрольные работы</p> <p>Промежуточная аттестация: Экзамен</p>

2. Критерии оценивания сформированности компетенций

Компетенция	Зачтено			Не зачтено
	Высокий уровень (отлично) (86-100 баллов)	Средний уровень (хорошо) (71-85 баллов)	Низкий уровень (удовлетворительно) (56-70 баллов)	Ниже порогового уровня (неудовлетворительно) (0-55 баллов)
ПК-8 Владение навыками использования различных технологий разработки программного обеспечения и систем искусственного интеллекта	Знает основные понятия и формулировки в области применения алгоритмов исследования данных, основные методы и паттерны применяемые для задач, связанных с получением информации из данных.	Знает основные понятия и формулировки в области применения алгоритмов исследования данных.	Ориентируется в основных понятиях и формулировках в области применения алгоритмов исследования данных.	Не знает основные понятия и формулировки в области применения алгоритмов исследования данных.
	Умеет осуществлять расчет на основании выбора алгоритма исследований данных, осуществлять расчет подготовленных и нормализованных данных	Умеет осуществлять расчет на основании выбора алгоритма исследований данных	Умеет осуществлять расчет на основании одного алгоритма исследований данных	Не умеет осуществлять расчет на основании выбора алгоритма исследований данных
	Владеет навыками применения алгоритмов исследования данных, владеет навыками расчета подготовленных и нормализованных данных	Владеет навыками применения алгоритмов исследования данных	Владеет навыками применения одного алгоритма исследования данных	Не владеет навыками применения алгоритмов исследования данных

3. Распределение оценок за формы текущего контроля и промежуточную аттестацию

5 семестр:

Текущий контроль:

Лабораторные работы – 30 баллов.

Контрольные работы – 20 баллов.

Итого: 30 баллов + 20 баллов = 50 баллов

Промежуточная аттестация – экзамен

Промежуточная аттестация состоит из оценочных средств одного типа – устный ответ. Вопросы промежуточной аттестации составляются на основе всех тем, которые были рассмотрены в течение семестра. Экзамен проходит по билетам. В каждом билете два вопроса.

Итого – 50 баллов

Общее количество баллов по дисциплине за текущий контроль и промежуточную аттестацию: 50+50=100 баллов.

Соответствие баллов и оценок:

86-100 – отлично

71-85 – хорошо

56-70 – удовлетворительно

0-55 – неудовлетворительно

4. Оценочные средства, порядок их применения и критерии оценивания

4.1. Оценочные средства текущего контроля

4.1.1. Лабораторные работы

Лабораторные работы представляют собой развернутые индивидуальные задания, целью выполнения которых является формирование практических навыков по реализации и применению алгоритмов по исследованию данных. Каждая лабораторная работа состоит из нескольких связанных друг с другом задач, последовательное выполнение которых приводит к получению финального результата. Проверяющим оценивается корректность выполнения задач, работоспособность полученного результата, а также полнота и обоснованность сопроводительных комментариев к результату.

4.1.1.2. Критерии оценивания

Баллы в интервале 86-100% от максимальных ставятся, если:

Все задачи лабораторной работы выполнены в полном объеме и без ошибок. Сопроводительное описание демонстрирует корректные и обоснованные выводы. Приведены развернутые ответы на поставленные вопросы.

Баллы в интервале 71-85% от максимальных ставятся, если:

Все задачи лабораторной работы выполнены в полном объеме и без грубых ошибок. Допускается наличие незначительных неточностей (не более трех). Сопроводительное описание демонстрирует корректные и обоснованные выводы. Приведены ответы на все поставленные вопросы.

Баллы в интервале 56-70% от максимальных ставятся, если:

Все задачи лабораторной работы выполнены в полном объеме. В работе присутствуют ошибки, которые не влияют на конечный результат. Сопроводительное описание демонстрирует корректные выводы. В работе присутствуют необоснованные выводы. Приведены ответы на все поставленные вопросы.

Баллы в интервале 0-55% от максимальных ставятся, если:

Лабораторная работа выполнена в неполном объеме. Студент не владеет теоретическим материалом, допуская грубые ошибки, испытывает затруднения в формулировке собственных суждений, не способен ответить на дополнительные вопросы.

4.1.1.3. Содержание оценочного средства

Образец лабораторной работы №1

Создать программный модуль, собирающий данные из сети Интернет и анализирующий текст с точки зрения разных факторов

Образец лабораторной работы №2

Создать программный модуль, собирающий данные по предложенному веб сайту из сети Интернет с последующим построением ранжирования страницы

Образец лабораторной работы №3

Создать программный модуль, использующий Фильтра Блума для организации обработки потоковых данных

4.1.2. Контрольная работа

4.1.2.1. Порядок проведения и процедура оценивания

В рамках курса предусмотрено 2 контрольных работы в формате разработки компьютерных программ по заданию.

4.1.2.2. Критерии оценивания

Баллы в интервале 86-100% от максимальных ставятся, если:

Правильно выполнены все задания. Продемонстрирован высокий уровень владения материалом. Проявлены превосходные способности применять знания и умения к выполнению конкретных заданий. В логических рассуждениях и обосновании решения нет пробелов и ошибок.

Баллы в интервале 71-85% от максимальных ставятся, если:

Правильно выполнена большая часть заданий. Присутствуют незначительные ошибки. Продемонстрирован хороший уровень владения материалом. Проявлены средние способности применять знания и умения к выполнению конкретных заданий.

Баллы в интервале 56-70% от максимальных ставятся, если:

Задания выполнены более чем наполовину. Присутствуют серьезные ошибки. Продемонстрирован удовлетворительный уровень владения материалом. Проявлены низкие способности применять знания и умения к выполнению конкретных заданий. Алгоритм решения реализован примерно наполовину, возможно, с вычислительными ошибками.

Баллы в интервале 0-55% от максимальных ставятся, если:

Задания выполнены менее чем наполовину. Продемонстрирован неудовлетворительный уровень владения материалом. Проявлены недостаточные способности применять знания и умения к выполнению конкретных заданий.

4.1.2.3. Содержание оценочного средства

Образец контрольной работы №1

На основании предложенного файла создать программный модуль реализующий анализ данных в выстраивании частотных пар по данным

Образец контрольной работы №2

В предложенных топологиях графовых зависимостей реализовать модуль ранжирования объектов

4.2. Оценочные средства промежуточной аттестации

4.2.1. Экзамен. Устный ответ

4.2.1.1. Порядок проведения

Вопросы промежуточной аттестации составляются на основе всех тем, которые были рассмотрены в течение семестра. Экзамен проходит по билетам. В каждом билете два вопроса из списка, приведенного в п.4.2.1.3.

4.2.1.2. Критерии оценивания

Баллы в интервале 86-100% от максимальных ставятся, если обучающийся:

Обнаружил всестороннее, систематическое и глубокое знание учебно-программного материала, умение свободно выполнять задания, предусмотренные программой, усвоил основную литературу и знаком с дополнительной литературой, рекомендованной программой дисциплины, усвоил взаимосвязь основных понятий дисциплины в их значении для приобретаемой профессии, проявил творческие способности в понимании, изложении и использовании учебно-программного материала.

Баллы в интервале 71-85% от максимальных ставятся, если обучающийся:

Обнаружил полное знание учебно-программного материала, успешно выполнил предусмотренные программой задания, усвоил основную литературу, рекомендованную программой дисциплины, показал систематический характер знаний по дисциплине и способен к их самостоятельному пополнению и обновлению в ходе дальнейшей учебной работы и профессиональной деятельности.

Баллы в интервале 56-70% от максимальных ставятся, если обучающийся:

Обнаружил знание основного учебно-программного материала в объеме, необходимом для дальнейшей учебы и предстоящей работы по профессии, справился с выполнением заданий, предусмотренных программой, знаком с основной литературой, рекомендованной программой дисциплины, допустил погрешности в ответе на экзамене и при выполнении экзаменационных заданий, но обладает необходимыми знаниями для их устранения под руководством преподавателя.

Баллы в интервале 0-55% от максимальных ставятся, если обучающийся:

Обнаружил значительные пробелы в знаниях основного учебно-программного материала, допустил принципиальные ошибки в выполнении предусмотренных программой заданий и не способен продолжить обучение или приступить по окончании университета к профессиональной деятельности без дополнительных занятий по соответствующей дисциплине.

4.2.1.3. Оценочные средства

Вопросы:

1. Наука Data Mining. Data Cleaning, Data Reduction, Data Transformation, Feature Selection
2. LSA / SAX методы
3. Типы данных в Data Mining
4. Similarity / Dissimilarity и Distances
5. Structure of Web, Early Search Engines
6. PageRank алгоритм. Matrix метод
7. PageRank алгоритм. Iteration метод
8. Ограничения PageRank.
9. Topic-Sensitive PageRank
10. Spam Farm.
11. TrustRank.
12. HITS алгоритм
13. Жадные алгоритмы
14. The Adwords Problem. Алгоритм Баланса
15. Frequent Itemsets. Apriori принцип
16. Frequent Itemsets. алгоритм PCY
17. Frequent Itemsets. Multistage алгоритм
18. Frequent Itemsets. Multihash алгоритм

19. Frequent Itemsets. SON алгоритм. Toivonen алгоритм
20. Bloom Filter.
21. Countable Bloom Filter.
22. Count-Distinct Problem. Flajolet-Martin алгоритм
23. The Alon-Matias-Szegedy алгоритм.
24. Counting Ones in a Window. Datar-Gionis-Indyk-Motwani алгоритм
25. Иерархическая кластеризация в Euclidean Space
26. Centroid (Clustroid), Radius and Diameter
27. Иерархическая кластеризация
28. K-means алгоритм
29. Алгоритм Bradley, Fayyad, Reina
30. CURE алгоритм
31. Алгоритм GRGPF
32. BDMO алгоритм

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 09.03.04 - Программная инженерия

Профиль подготовки: Искусственный интеллект и современная разработка программного обеспечения

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: английский

Год начала обучения по образовательной программе: 2025

Основная литература:

1. Лимановская, О. В. Основы машинного обучения : учебное пособие / О. В. Лимановская, Т. И. Алферьева. - 2-е изд., стер. - Москва : ФЛИНТА : Издательство Уральского университета, 2022. - 88 с. - ISBN 978-5-9765-5006-3 (ФЛИНТА); ISBN 978-5-7996-3015-7 (Изд-во Урал. ун-та). - Текст : электронный. - URL: <https://znanium.com/catalog/product/1891377> (дата обращения: 12.12.2022). – Режим доступа: по подписке.
2. Флах П., Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных: учебное пособие / Флах П. - Москва: ДМК Пресс, 2015. - 400 с. - ISBN 978-5-97060-273-7 - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <http://www.studentlibrary.ru/book/ISBN9785970602737.html> (дата обращения: 12.12.2022). - Режим доступа : по подписке.
3. Рашка С., Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения: практическое пособие / Рашка С. - Москва: ДМК Пресс, 2017. - 418 с. - ISBN 978-5-97060-409-0 - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <http://www.studentlibrary.ru/book/ISBN9785970604090.html> (дата обращения: 12.12.2022). - Режим доступа : по подписке.

Дополнительная литература:

1. Макшанов, А. В. Технологии интеллектуального анализа данных : учебное пособие / А. В. Макшанов, А. Е. Журавлев. — 2-е изд., стер. — Санкт-Петербург : Лань, 2022. — 212 с. — ISBN 978-5-8114-4493-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/206711> (дата обращения: 12.12.2022). — Режим доступа: для авториз. пользователей.
2. Теория вероятностей и математическая статистика : учебное пособие / Л.Г. Бирюкова, Г.И. Бобрик, Р.В. Сагитов [и др.] ; под ред. В.И. Матвеева. — 2-е изд., испр. и доп. — Москва : ИНФРА-М, 2021. — 289 с. — (Высшее образование: Бакалавриат). — DOI 10.12737/18865. - ISBN 978-5-16-011793-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1237088> (дата обращения: 12.12.2022). - Режим доступа : по подписке.
3. Зарова, Е. В. Методы Data mining в обработке и анализе статистических данных (решения в R) : монография / Е.В. Зарова. — Москва : ИНФРА-М, 2021. — 232 с. : ил. - ISBN 978-5-16-016814-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1240276> (дата обращения: 12.12.2022). – Режим доступа: по подписке.

Перечень информационных технологий, включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 09.03.04 - Программная инженерия

Профиль подготовки: Искусственный интеллект и современная разработка программного обеспечения

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: английский

Год начала обучения по образовательной программе: 2025

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

1. Операционная система Microsoft office professional plus 2010, или Microsoft Windows 7 Профессиональная
2. Пакет офисного программного обеспечения Microsoft Office 365, или Microsoft office professional plus 2010
3. Adobe Reader XI или Adobe Acrobat Reader DC
4. Браузер Mozilla Firefox
5. Браузер Google Chrome
6. Kaspersky Endpoint Security для Windows
7. Электронная библиотечная система «ZNANIUM.COM»
8. Электронная библиотечная система Издательства «Лань»
9. Электронная библиотечная система «Консультант студента»