

**Министерство науки и высшего образования РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Национальный исследовательский университет «МЭИ»**

Направление подготовки/специальность: 27.03.04 Управление в технических системах

Наименование образовательной программы: Управление и информатика в технических системах

Уровень образования: высшее образование - бакалавриат

Форма обучения: Очная


**Оценочные материалы
по дисциплине
Прикладной анализ данных**

**Москва
2022**

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ РАЗРАБОТАЛ:

Преподаватель

(должность)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Толчеев В.О.
	Идентификатор	Rfbd680da-TolcheevVO-692f9924

(подпись)


В.О. Толчеев

(расшифровка
подписи)

СОГЛАСОВАНО:

Руководитель
образовательной
программы

(должность, ученая степень, ученое
звание)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Сидорова Е.Ю.
	Идентификатор	R0dee6ce9-SidorovaYY-923dc6a8

(подпись)


Е.Ю.

Сидорова

(расшифровка
подписи)

Заведующий
выпускающей кафедры

(должность, ученая степень, ученое
звание)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Бобряков А.В.
	Идентификатор	R2c90f415-BobriakovAV-70dec1fa

(подпись)

А.В.

Бобряков

(расшифровка
подписи)

ОБЩАЯ ЧАСТЬ

Оценочные материалы по дисциплине предназначены для оценки: достижения обучающимися запланированных результатов обучения по дисциплине, этапа формирования запланированных компетенций и уровня освоения дисциплины.

Оценочные материалы по дисциплине включают оценочные средства для проведения мероприятий текущего контроля успеваемости и промежуточной аттестации.

Формируемые у обучающегося компетенции:

1. ПК-2 Способен разрабатывать и применять технологии сбора, обработки и анализа разнотипных данных для расчета и проектирования систем и средств автоматизации и управления

ИД-1 Может формировать выборки и подготавливать данные для проведения анализа

ИД-4 Использует стандартное программное обеспечение и специализированные библиотеки для обработки и анализа данных

и включает:

для текущего контроля успеваемости:

Форма реализации: Защита задания

1. Проведение контрольной работы №1 (Лабораторная работа)

2. Проведение контрольной работы №2 (Лабораторная работа)

3. Проведение контрольной работы №3 (Лабораторная работа)

4. Проведение контрольной работы №4 (Лабораторная работа)

БРС дисциплины

6 семестр

Раздел дисциплины	Веса контрольных мероприятий, %				
	Индекс КМ:	КМ-1	КМ-2	КМ-3	КМ-4
	Срок КМ:	4	6	12	14
Электронные информационные ресурсы в области науки и техники					
Информационные ресурсы со свободным доступом	+	+	+		
Информационный поиск и рекомендательные системы	+	+	+		
Методика поиска, обработки и анализа разнотипных данных					
Выбор подходящих моделей и методов анализа в зависимости от решаемой задачи				+	+
Программные средства анализа данных				+	+
	Вес КМ:	20	20	20	40

§Общая часть/Для промежуточной аттестации§

СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ТЕКУЩЕГО КОНТРОЛЯ

I. Оценочные средства для оценки запланированных результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Индекс компетенции	Индикатор	Запланированные результаты обучения по дисциплине	Контрольная точка
ПК-2	ИД-1 _{ПК-2} Может формировать выборки и подготавливать данные для проведения анализа	Знать: основы информационного поиска и анализа научной информации, способы повышения релевантности и пертинентности, основные электронные ресурсы в области Data Science Уметь: формировать выборки из электронных научных библиотек, коллекций фактографических и мультимедийных данных	Проведение контрольной работы №1 (Лабораторная работа) Проведение контрольной работы №2 (Лабораторная работа)
ПК-2	ИД-4 _{ПК-2} Использует стандартное программное обеспечение и специализированные библиотеки для обработки и анализа данных	Знать: методы обработки, визуализации и анализа разнотипных многомерных данных Уметь: применять методы интеллектуального анализа, реализованные в стандартных библиотеках	Проведение контрольной работы №3 (Лабораторная работа) Проведение контрольной работы №4 (Лабораторная работа)

		Python и общедоступных программных средствах, для решения прикладных задач	
--	--	--	--

II. Содержание оценочных средств. Шкала и критерии оценивания

КМ-1. Проведение контрольной работы №1

Формы реализации: Защита задания

Тип контрольного мероприятия: Лабораторная работа

Вес контрольного мероприятия в БРС: 20

Процедура проведения контрольного мероприятия: Выдаются вопросы по изученным темам (по вариантам). Время выполнения – 30 минут. Ответ должен содержать краткие ответы на вопросы.

Краткое содержание задания:

Перечислить и дать краткую характеристику электронным информационным ресурсам в области науки и образования. Провести сравнительный анализ, указать области применения, привести определение наукометрических показателей и способов их расчета.

Контрольные вопросы/задания:

Уметь: формировать выборки из электронных научных библиотек, коллекций фактографических и мультимедийных данных	1. Как рассчитывается индекс Хирша и что он показывает? 2. Как определить импакт-фактор журнала?
---	---

Описание шкалы оценивания:

Оценка: 5

Нижний порог выполнения задания в процентах: 85

Описание характеристики выполнения знания: Оценка "отлично" выставляется если задание выполнено в полном объеме или выполнено преимущественно верно.

Оценка: 4

Нижний порог выполнения задания в процентах: 65

Описание характеристики выполнения знания: Оценка "хорошо" выставляется если большинство вопросов раскрыто. выбрано верное направление для решения задач.

Оценка: 3

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется если задание преимущественно выполнено.

КМ-2. Проведение контрольной работы №2

Формы реализации: Защита задания

Тип контрольного мероприятия: Лабораторная работа

Вес контрольного мероприятия в БРС: 20

Процедура проведения контрольного мероприятия: Выдаются два вопроса по изученным темам (по вариантам). Первый вопрос – теоретический. Второй вопрос – практический (например, осуществить поиск и анализ информации в открытой библиотеке eLibrary). Время выполнения – 45 минут. Ответ должен содержать краткую теоретическую часть и описание результатов выполнения практической части.

Краткое содержание задания:

В соответствии с заданной информационной потребностью пользователя сформировать осуществить поиск и сформировать выборку публикаций из цифровой библиотеки eLibrary. Оценить качество поиска. Сформулировать предложения по улучшению качества поиска и повышению релевантности выдачи.

Контрольные вопросы/задания:

Знать: основы информационного поиска и анализа научной информации, способы повышения релевантности и пертинентности, основные электронные ресурсы в области Data Science	1.Что такое качество поиска? 2.Что такое релевантность и пертинентность?
--	---

Описание шкалы оценивания:

Оценка: 5

Нижний порог выполнения задания в процентах: 85

Описание характеристики выполнения знания: Оценка "отлично" выставляется если задание выполнено в полном объеме или выполнено преимущественно верно.

Оценка: 4

Нижний порог выполнения задания в процентах: 65

Описание характеристики выполнения знания: Оценка "хорошо" выставляется если большинство вопросов раскрыто. выбрано верное направление для решения задач.

Оценка: 3

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется если задание преимущественно выполнено.

КМ-3. Проведение контрольной работы №3

Формы реализации: Защита задания

Тип контрольного мероприятия: Лабораторная работа

Вес контрольного мероприятия в БРС: 20

Процедура проведения контрольного мероприятия: Выдается задание по построению «типовой» рекомендательной системы – РС (по вариантам). В зависимости от области применения и специфики задания необходимо выбрать и обосновать способ реализации РС.

Краткое содержание задания:

Проанализировать специфику задачи. Обосновать способ построения РС, нарисовать структурную схему и описать процедуру функционирования РС. Проанализировать возможные подходы к улучшению качества работы РС.

Контрольные вопросы/задания:

Уметь: применять методы интеллектуального анализа, реализованные в стандартных библиотеках Python и общедоступных программных средствах, для решения прикладных задач	1.Приведите пример использования метода к-ближайших соседей в РС? 2.Как определять наличие дубликатов (нечетких дубликатов) в случае обработки текстовых данных?
---	---

Описание шкалы оценивания:

Оценка: 5

Нижний порог выполнения задания в процентах: 85

Описание характеристики выполнения знания: Оценка "отлично" выставляется если задание выполнено в полном объеме или выполнено преимущественно верно.

Оценка: 4

Нижний порог выполнения задания в процентах: 65

Описание характеристики выполнения знания: Оценка "хорошо" выставляется если большинство вопросов раскрыто. выбрано верное направление для решения задач.

Оценка: 3

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется если задание преимущественно выполнено.

КМ-4. Проведение контрольной работы №4

Формы реализации: Защита задания

Тип контрольного мероприятия: Лабораторная работа

Вес контрольного мероприятия в БРС: 40

Процедура проведения контрольного мероприятия: Выдается датасет (или по желанию студента может быть продолжена работа из предыдущего теста по построению РС). Проводится его анализ и обосновывается, какую модель (и каким методом) целесообразно построить. Применяются свободно распространяемые программные средства и библиотеки. Дается интерпретация полученных результатов и анализируются способы их улучшения.

Краткое содержание задания:

Проанализировать особенности датасета, выявить информативные признаки. Выбрать и обосновать методы построения модели. Провести сравнительный анализ моделей, оценить их качество. Подготовить краткий отчет о проведенных исследованиях.

Контрольные вопросы/задания:

Знать: методы обработки, визуализации и анализа разнотипных многомерных данных	1.Что такое сэмплирование и когда применяется? 2.Как определяется класс наблюдения в случае использования логистической регрессии при многоклассовой классификации?
--	--

Описание шкалы оценивания:

Оценка: 5

Нижний порог выполнения задания в процентах: 85

Описание характеристики выполнения знания: Оценка "отлично" выставляется если задание выполнено в полном объеме или выполнено преимущественно верно.

Оценка: 4

Нижний порог выполнения задания в процентах: 65

Описание характеристики выполнения знания: Оценка "хорошо" выставляется если большинство вопросов раскрыто. выбрано верное направление для решения задач.

Оценка: 3

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется если задание преимущественно выполнено.

СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

6 семестр

Форма промежуточной аттестации: Зачет

Пример билета

Наукометрический анализ. Виды рекомендательных систем.

Процедура проведения

Зачет без оценки проводится в устной форме по билетам. На подготовку ответа студенту отводится 45 минут.

1. Перечень компетенций/индикаторов и контрольных вопросов проверки результатов освоения дисциплины

1. Компетенция/Индикатор: ИД-1_{ПК-2} Может формировать выборки и подготавливать данные для проведения анализа

Вопросы, задания

1. Понятие информационной потребности пользователя.
2. Релевантность и пертинентность.
3. Обратная связь по релевантности.
4. Научные рекомендательные системы. Особенности реализации.

Материалы для проверки остаточных знаний

1. В чем заключается обратная связь по релевантности (метод Роккио).
Верный ответ: Метод Роккио позволяет модифицировать исходный поисковый запрос. Для этого пользователю предлагается указать документы, которые соответствуют его информационным потребностям (множество релевантных документов). Затем составляется модифицированный запрос, в котором термины, вошедшие во множество релевантных документов, увеличивают свои веса, а веса терминов из множества нерелевантных документов уменьшаются (или остаются неизменными). Для успешной работы метода Роккио надо правильно задать специальные коэффициенты
2. Какие существуют рекомендательные системы (РекС).
Верный ответ: 1) РекС на основе контентной фильтрации (Content-based filtering), 2) РекС на основе коллаборативной фильтрации (Collaborative filtering) и гибридные рекомендательные системы (Hybrid recommender system), использующие оба подхода.
3. Перечислите основные алгоритмы предварительной обработки текстовых документов.
Верный ответ: 1. Очистка слов от знаков препинания и ненужных символов; 2. Приведение текста к одному регистру; 3. Удаление стоп-слов. 4. Стемминг – процесс нахождения основы слова для заданного исходного слова. При этом основа может не совпадать с морфологическим корнем слова; 5. Лемматизация – это процесс с использованием лексикона и морфологического анализа слов, в результате которого удаляются только флективные окончания, и возвращается основная форма слова, называемая леммой.
4. Как рассчитывается и зачем используется TF-IDF – взвешивание?
Верный ответ: TF-IDF взвешивание присваивает вес слову i в документе j пропорционально числу вхождений слова в данный документ и обратно

пропорционально общему числу документов в выборке, в которых также содержится это слово. Такой подход позволяет снизить веса общеупотребительных высокочастотных слов.

5. Что такое релевантность и пертинентность?

Верный ответ: Релевантность — соответствие поискового намерения, заложенного (сформулированного) в запросе, и выдачи в информационно-поисковой системе, полученной в результате этого запроса. Пертинентность — соответствие найденных информационно-поисковой системой результатов информационным потребностям пользователя, независимо от того, как полно и как точно эта информационная потребность выражена в тексте информационного запроса.

6. Что такое квартиль научного журнала?

Верный ответ: Квартиль научного журнала — это категория научных журналов, которая определяется в зависимости от цитируемости журнала, показывая востребованность журнала научным сообществом. В первую и вторую квартили входят 50 % журналов, индексируемых в соответствующей международной базе данных, имеющие наибольшие библиометрические показатели.

7. Из чего состоит библиографическое описание (БО) документа?

Верный ответ: БО состоит из следующих основных частей: 1) Название статьи. 2) Список авторов. 3) Аннотация. Самая большая и содержательная часть БО. Обычно включает в себя около 3-4 предложений, в которых даётся краткое описание всей статьи. 4) Ключевые слова. 5) Дата и место публикации. Также библиографическое описание может содержать различную второстепенную информацию: размер исходной статьи, ее цитирование и др.

2. Компетенция/Индикатор: ИД-4ПК-2 Использует стандартное программное обеспечение и специализированные библиотеки для обработки и анализа данных

Вопросы, задания

1. Критерии качества бинарной и многоклассовой классификации.
2. Сбалансированные и несбалансированные выборки.
3. Методы сэмплирования данных.

Материалы для проверки остаточных знаний

1. Укажите, какие методы позволяют решать задачу регрессии:

Ответы:

- 1 - метод наименьших квадратов,
- 2 - метод k-ближайших соседей,
- 3 - деревья решений,
- 4 - логистическая регрессия,
- 5 - дискриминантный анализ,
- 6 - метод наименьших модулей,
- 7 - случайный лес.

Верный ответ: Правильный ответ: 1,2,3,6,7.

2. Укажите, какие методы позволяют решать задачу классификации:

Ответы:

- 1 - метод наименьших квадратов,
- 2 - метод k-ближайших соседей,
- 3 - деревья решений,
- 4 - логистическая регрессия,
- 5 - дискриминантный анализ,
- 6 - метод наименьших модулей,
- 7 - случайный лес.

Верный ответ: Правильный ответ: 2,3,4,5,7.

3. Укажите, какие методы позволяют решать задачу кластеризации:

Ответы:

- 1 - метод наименьших квадратов,
- 2 - метод k-средних,
- 3 – метод DBSCAN,
- 4 - логистическая регрессия,
- 5 - дискриминантный анализ,
- 6 - метод наименьших модулей,
- 7 – случайный лес.

Верный ответ: Правильный ответ: 2,3,7.

4. Какие методы используются для выявления дубликатов (одинаковых документов) и почти дубликатов?

Ответы:

- 1 - метод наименьших квадратов,
- 2 - метод k-средних,
- 3 – коэффициент ассоциативности Жаккара,
- 4 - логистическая регрессия,
- 5 – метод Джаро-Винклера,
- 6 - метод наименьших модулей,
- 7 – метод шинглов.

Верный ответ: Ответ: 3,5,7.

5. Дайте определение полного и нечеткого (почти) дубликата текстового документа.

Верный ответ: Дубликат текстового документа – это документ идентичный исходному. Идентичность может быть как лексическая (определяется набором терминов), так и смысловая (может быть определена только экспертно). С лексической точки зрения выделяют: Полные дубликаты – полностью совпадающие (имеющие одинаковый набор терминов) документы. Нечёткие (неполные) дубликаты (почти дубликаты) – документы, имеющие близкое лексическое содержание (т.е. использующие практически один и тот же набор терминов)

6. Для расчета F1-меры необходимо знать:

Ответы:

- Значение точности (*precision*);
- Значение полноты (*recall*);
- Значение доли верных ответов (*accuracy*).

Верный ответ: *Значение точности (*precision*); *Значение полноты (*recall*); Значение доли верных ответов (*accuracy*).

7. В каких случаях не рекомендуется использовать показатель *Accuracy* (доля верных ответов) для оценки качества классификации?

Ответы:

- 1 – при сбалансированных классах;
- 2 – при несбалансированных классах;
- 3 – при бинарной классификации;
- 4 – при многоклассовой классификации.

Верный ответ: Ответ: 2.

II. Описание шкалы оценивания

Оценка: 5

Нижний порог выполнения задания в процентах: 85

Описание характеристики выполнения знания: Работа выполнена в рамках "продвинутого" уровня. Ответы даны верно, четко сформулированные особенности практических решений.

Оценка: 4

Нижний порог выполнения задания в процентах: 65

Описание характеристики выполнения знания: Работа выполнена в рамках "базового" уровня. Большинство ответов даны верно. В части материала есть незначительные недостатки.

Оценка: 3

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Работа выполнена в рамках "порогового" уровня. Основная часть задания выполнена верно.

III. Правила выставления итоговой оценки по курсу

Положением о балльно-рейтинговой системе для студентов НИУ «МЭИ» на основании семестровой и аттестационной составляющих.