

МИНОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ

Директор института прикладной
математики и компьютерных наук

А.В. Замятин

« 11 » _____ 2021 г.



Фонд оценочных средств по дисциплине

Анализ больших массивов данных (Big Data)

Направление подготовки

01.03.02 Прикладная математика и информатика

Профиль: **Прикладная математика и информатика**

ОС составил:

д-р техн. наук, профессор,

заведующий кафедрой теоретических основ информатики


А.В. Замятин

Рецензент:

д-р техн. наук, профессор,

заведующий кафедрой прикладной информатики


С.П. Сущенко

Фонд оценочных средств одобрен на заседании учебно-методической комиссии института прикладной математики и компьютерных наук (УМК ИПМКН).

Протокол от 17 июня 2021 г. № 05

Председатель УМК ИПМКН,

д-р техн. наук, профессор


С.П. Сущенко

Оценочные средства (ОС) являются элементом системы оценивания сформированности компетенций у обучающихся в целом или на определенном этапе ее формирования.

ОС разрабатывается в соответствии с рабочей программой (РП) дисциплины.

1. Компетенции и результаты обучения, формируемые в результате освоения дисциплины

Компетенция	Индикатор компетенции	Код и наименование результатов обучения (планируемые результаты обучения, характеризующие этапы формирования компетенций)	Критерии оценивания результатов обучения	
			Зачтено	Не зачтено
ОПК-2 – Способен использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач.	ИОПК-2.1. Обладает навыками объектно-ориентированного программирования для решения прикладных задач в профессиональной деятельности.	ОР-2.1.1. Уметь применять полученные знания объектно-ориентированного программирования при разработке программ	Сформированные систематические знания и умения, возможно содержащие отдельные пробелы знания и умения	Отсутствие знаний и умений
	ИОПК-2.2. Проявляет навыки использования основных языков программирования, основных методов разработки программ, стандартов оформления программной документации.	ОР-2.2.1. Знать языки программирования C# и Python. ОР-2.2.2. Знать библиотеки numpy, pandas, matplotlib для работы с искусственным интеллектом на языке Python. ОР-2.2.3. Уметь работать с online компиляторами как средствами редактирования, отладки, компиляции и выполнения программ.	Сформированные систематические знания и умения, возможно содержащие отдельные пробелы знания и умения	Отсутствие знаний и умений

	ИОПК-2.3. Демонстрирует умение отбора среди существующих математических методов, наиболее подходящих для решения конкретной прикладной задачи.	ОР-2.3.1. Знает основные методы численного анализа математических моделей и границы применимости данных методов. ОР-2.3.2. Умеет осуществлять выбор метода численного решения задачи математического моделирования. ОР-2.3.3. Владеет цифровыми инструментами, необходимыми для численного анализа математических моделей при решении конкретной прикладной задачи	Сформированные систематические знания и умения, возможно содержащие отдельные пробелы знания и умения	Отсутствие знаний и умений
	ИОПК-2.4. Демонстрирует умение адаптировать существующие математические методы для решения конкретной прикладной задачи.	ОР-2.4.1. Владеет опытом адаптации существующих математических методов для решения конкретной прикладной задачи	Сформированные систематические знания и умения, возможно содержащие отдельные пробелы знания и умения	Отсутствие знаний и умений
ОПК-4 – Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности.	ИОПК-4.1. Обладает необходимыми знаниями в области информационных технологий, в том числе понимает принципы их работы.	ОР-4.1.1. Знать принципы работы, используемые в информационных технологиях	Сформированные систематические знания и умения, возможно содержащие отдельные пробелы знания и умения	Отсутствие знаний и умений
	ИОПК-4.2. Применяет знания, полученные в области информационных технологий, при решении задач профессиональной деятельности.	ОР-4.2.1. Уметь применять знания, полученные в области информационных технологий, при решении задач профессиональной деятельности	Сформированные систематические знания и умения, возможно содержащие отдельные пробелы знания и умения	Отсутствие знаний и умений

ПК-3	<p>ИПК-3.1. Реализовывает построение формализованной математической модели системы (подсистемы), введение целевой функции системы, подсистемы и ограничений, соответствующих требованиям к системе (подсистеме).</p>	<p>ОР-3.1.1: Знать основные методы построения формализованной математической модели в задачах интеллектуального анализа данных и других областях с использованием информационных технологий. ОР-3.1.2: Знать существующие методы и подходы к интеллектуальному анализу данных различной природы. ОР-3.1.3: Уметь применять существующие методы интеллектуального анализа данных, обоснованно адаптируя и модифицируя их с учетом особенностей задачи предметной области. ОР-3.1.4: Уметь формулировать научно-практическую задачу, планировать ее решение и выполнить в соответствии с планом.</p>	<p>Сформированные систематические знания и умения, возможно содержащие отдельные пробелы знания и умения</p>	<p>Отсутствие знаний и умений</p>
	<p>ИПК-3.2. Адаптирует формализованную математическую модель системы (подсистемы) к изменению требований (ограничений к целевой функции) к системе (подсистеме).</p>	<p>ОР-3.2.1. Научиться адаптировать формализованную математическую модель в задачах интеллектуального анализа данных. ОР-3.2.2. Научиться использовать библиотеки для работы с большими данными и искусственным интеллектом.</p>	<p>Сформированные систематические знания и умения, возможно содержащие отдельные пробелы знания и умения</p>	<p>Отсутствие знаний и умений</p>

	<p>ИПК-3.3. Выявляет и формализовывает в виде математической модели возникающие при функционировании системы (подсистемы) риски; выявляет и анализирует проблемные ситуации.</p>	<p>ОП-3.3.1. Уметь применять на практике структуры данных для хранения и обработки данных.</p>	<p>Сформированные систематические умения, возможно содержащие отдельные пробелы</p>	<p>Отсутствие умений</p>
--	--	--	---	--------------------------

2. Этапы формирования компетенций и виды оценочных средств

№	Этапы формирования компетенций (разделы дисциплины)	Код и наименование результатов обучения	Вид оценочного средства (тесты, задания, кейсы, вопросы и др.)
1.	Раздел 1. Основные проблемы построения систем	ОР-2.1.1, ОР-2.2.1, ОР-2.2.2, ОР-2.2.3, ОР-2.3.1, ОР-2.3.2, ОР-2.3.3, ОР-2.4.1, ОР-3.1.1, ОР-3.1.2, ОР-3.1.3, ОР-3.1.4, ОР-3.2.1, ОР-3,2.2, ОР-3.3.1.	Задания для текущего контроля, вопросы для проведения промежуточной аттестации
2.	Раздел 2. Предварительная обработка данных. Классификация.	ОР-2.1.1, ОР-2.2.1, ОР-2.2.2, ОР-2.2.3, ОР-2.3.1, ОР-2.3.2, ОР-2.3.3, ОР-2.4.1, ОР-3.1.1, ОР-3.1.2, ОР-3.1.3, ОР-3.1.4, ОР-3.2.1, ОР-3,2.2, ОР-3.3.1.	Задания для текущего контроля, вопросы для проведения промежуточной аттестации
3.	Раздел 3. Регрессия. Ассоциация, последовательная ассоциация, аномалии и визуализация.	ОР-2.1.1, ОР-2.2.1, ОР-2.2.2, ОР-2.2.3, ОР-2.3.1, ОР-2.3.2, ОР-2.3.3, ОР-2.4.1, ОР-3.1.1, ОР-3.1.2, ОР-3.1.3, ОР-3.1.4, ОР-3.2.1, ОР-3,2.2, ОР-3.3.1.	Задания для текущего контроля, вопросы для проведения промежуточной аттестации
4.	Раздел 4. Высокопроизводительная обработка данных. Программные среды для интеллектуального анализа данных.	ОР-2.1.1, ОР-2.2.1, ОР-2.2.2, ОР-2.2.3, ОР-2.3.1, ОР-2.3.2, ОР-2.3.3, ОР-2.4.1, ОР-3.1.1, ОР-3.1.2, ОР-3.1.3, ОР-3.1.4, ОР-3.2.1, ОР-3,2.2, ОР-3.3.1.	Задания для текущего контроля, вопросы для проведения промежуточной аттестации

3. Типовые контрольные задания или иные материалы, необходимые для оценки образовательных результатов обучения

3.1. Типовые задания для проведения текущего контроля успеваемости по дисциплине

Реферат (на согласованную тему). К реферату необходимо сделать презентацию.

Примеры тем:

Современные нейронные сети в обработке данных (изображений, видео, технологических сигналов, музыки и т.п.);

Современные алгоритмы классификации (изображений, текстов и т.п.);

Интеллектуальная обработка данных в ... (промышленности, медицине, бизнесе, индустрии развлечений, досуга и др.);

Извлечение знаний из текстов;

Детектирование аномалий;

Разновидности сверточных нейронных сетей;

Интеллектуальные алгоритмы в ранней диагностике заболеваний;

Интеллектуальные алгоритмы в персонализированной медицине;

Интеллектуальные алгоритмы в робототехнике, транспортных системах и т.п.;

Интеллектуальные алгоритмы в банковском деле/страховании/...;

Проект (на согласованную тему). Реализовать небольшой проект по интеллектуальной обработке данных с использованием среды RapidMiner или одного из языков программирования (например, Python, R), с возможным использованием общедоступных баз данных (или данных из иных источников).

Этапы реализации проекта:

Поиск и подготовка набора данных;

Разработка технического задания;

Пилотная реализация одной модели, выбор метрики и оценка точности (фиксация полученной точности на этом этапе);

Реализация всех пунктов технического задания, настройка параметров моделей, оценка точности (точность, полученная на этом этапе должна быть больше чем на предыдущем):

Подготовка отчета (с описанием предметной области, выбранных алгоритмов и параметров моделей), презентации, публичная защита проекта;

Каждый студент реализует индивидуальный или групповой проект как последовательность лабораторных работ:

Лабораторная работа №1. Индивидуальное задание по теме «Анализ предметной области, формулировка целей и задач исследования. Извлечение и первичное сохранение данных».

Цель работы – научить студентов решать задачи анализа предметной области, ее адаптации для методов анализа данных с учетом принципиальных особенностей предметной области.

Лабораторная работа №2. Индивидуальное задание по теме «Предварительная обработка данных: очистка, интеграция, преобразование».

Цель работы – научить студентов решать задачи предварительной обработки данных, предполагающей трудоемкую процедуру очистки (исключение противоречий, случайных выбросов и помех, пропусков), интеграции (объединение данных из нескольких возможных источников в одном хранилище), преобразования (может включать агрегирование и сжатие данных, дискретизацию атрибутов и сокращение размерности и т.п.).

Лабораторная работа №3. Индивидуальное задание по теме «Содержательный анализ данных методами Data Mining».

Цель работы – научить студентов обоснованно применять базовые методы интеллектуального анализа данных, учитывая особенности как теоретического построения применяемых методов, так и выбранной предметной области.

Лабораторная работа №4. Индивидуальное задание по теме «Визуализация и интерпретация полученных результатов».

Цель работы – научить студентов выполнять визуализацию и интерпретация полученных результатов в виде, пригодном для принятия управленческих решений.

Примеры тем для самостоятельного изучения:

- Нейросетевые методы анализа данных, сверточные сети (convolution neural networks), глубинное обучение (deep learning).
- Методы интеллектуального анализа медиа (social media data mining).
- Методы машинного обучения в задачах финансовой аналитики.
- Методы машинного обучения в задачах ранней медицинской диагностики.
- Комбинирование моделей в анализе данных, бустинг.
- Метод анализа независимых компонент (independent component analysis).
- Методы визуализации данных высокой размерности.

3.2. Типовые задания для проведения промежуточной аттестации по дисциплине

Вопросы к зачету:

1. Основные понятия, терминология;
2. Data Mining / Data Science;
3. Big Data (основные понятия и свойства);
4. Дедукция и индукция;
5. Интеллектуальный анализ данных в бизнесе примеры применения;
6. Интеллектуальный анализ данных в решении сложных прикладных задач;
7. Интеллектуальный анализ данных в ранней диагностике опасных заболеваний;
8. Интеллектуальный анализ данных в индустриальной предиктивной аналитике;
9. Основные задачи и классификация методов анализа данных;
10. Принципиальные основы машинного обучения;
11. Предварительная обработка данных;
12. Оптимизация признакового пространства;
13. Постановка задачи классификации;
14. Контролируемая непараметрическая классификация;
15. Контролируемая непараметрическая нейросетевая классификация;
16. Классификация по методу машины опорных векторов;
17. Деревья решений;
18. Неконтролируемая классификация (кластеризация);
19. Регрессия (понятие регрессии, основные этапы регрессионного анализа, методы восстановления регрессии);
20. Ассоциация;
21. Последовательная ассоциация (алгоритмы семейства «Априори», алгоритм GSP);
22. Многоуровневое машинное обучение (бутстрэппинг, бэггинг, стекинг, бустинг);
23. Обнаружение аномалий;
24. Визуализация в Data Mining;
25. Функции активации;
26. Основные типы искусственных нейронных сетей;
27. Сверточные нейронные сети;
28. Среды и фреймворки глубинного обучения;
29. Основные задачи обработки текста;
30. Этапы предварительной обработки текста;
31. Метрики качества классификации;
32. Гипотеза A/B, Каппа-индекс согласия, ROC-кривая;
33. Метрика качества прогноза временного ряда;
34. Метрики качества кластеризации;

35. Принципы высокопроизводительных вычислений;
36. Особенности построения вычислительного кластера;
37. Среды и инструменты высокопроизводительных вычислений;
38. Инструменты data mining.

4. Методические материалы, определяющие процедуры оценивания образовательных результатов обучения

4.1. Методические материалы для оценки текущего контроля успеваемости по дисциплине.

Рейтинговая система для оценки текущей успеваемости обучающихся

Таблица – Балльные оценки для элементов контроля

Элементы учебной деятельности	Максимальный балл с начала семестра	Оцениваемая компетенция
Реферат по теме с презентацией	20	ОПК-2, ОПК-4, ПК-3
Реализация проекта	40	ОПК-2, ОПК-4, ПК-3
Опрос на занятиях	10	ОПК-2, ОПК-4, ПК-3
Зачет с оценкой	30	ОПК-2, ОПК-4, ПК-3

4.2. Методические материалы для проведения промежуточной аттестации по дисциплине.

Сумма баллов, набранная студентом в течение семестра, переводится в оценку промежуточной аттестации успеваемости студента по приведенной ниже шкале.

Пересчет баллов в оценки для промежуточной аттестации

Баллы на дату контрольной точки	Оценка
$\geq 65\%$ от максимальной суммы баллов	зачтено
$< 65\%$ от максимальной суммы баллов	не зачтено