МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ
УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ»
(ТУСУР)

УТВЕРЖДАЮ
Проректор по УР
_____Сенченко П.В.
«22» _____02_____ 2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

НАБОРЫ ДАННЫХ В ЗАДАЧАХ МАШИННОГО ОБУЧЕНИЯ

Уровень образования: **высшее образование - магистратура**
Направление подготовки / специальность: **09.04.04 Программная инженерия**
Направленность (профиль) / специализация: **Искусственный интеллект в безопасности киберфизических систем**
Форма обучения: **очная**
Факультет: **Факультет безопасности (ФБ)**
Кафедра: **Кафедра комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС)**
Курс: **1**
Семестр: **1**
Учебный план набора 2023 года

Объем дисциплины и виды учебной деятельности

Виды учебной деятельности	1 семестр	Всего	Единицы
Лекционные занятия	8	8	часов
Лабораторные занятия	20	20	часов
в т.ч. в форме практической подготовки	12	12	часов
Самостоятельная работа	116	116	часов
Общая трудоемкость (включая промежуточную аттестацию)	144	144	часов
	4	4	з.е.

Формы промежуточной аттестация	Семестр
Зачет	1

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Сенченко П.В.
Должность: Проректор по УР
Дата подписания: 22.02.2023
Уникальный программный ключ:
a1119608-cdff-4455-b54e-5235117c185c

Томск

Согласована на портале № 74823

1. Общие положения

1.1. Цели дисциплины

1. Формирование фундаментальных знаний и практических умений в области обработки данных для решения задач машинного обучения.

1.2. Задачи дисциплины

1. Развитие навыков по поиску, преодолению и исправлению проблем анализируемых данных, необходимых для решения прикладных задач в сфере анализа данных.
2. Получение знаний о нормативно-правовых аспектах публикации и использования данных.

2. Место дисциплины в структуре ОПОП

Блок дисциплин: Б1. Дисциплины (модули).
Часть блока дисциплин: Часть, формируемая участниками образовательных отношений.
Модуль дисциплин: Модуль профессиональной подготовки (major).
Индекс дисциплины: Б1.В.01.02.
Реализуется с применением электронного обучения, дистанционных образовательных технологий.

3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Процесс изучения дисциплины направлен на формирование следующих компетенций в соответствии с ФГОС ВО и основной образовательной программой (таблица 3.1):
Таблица 3.1 – Компетенции и индикаторы их достижения

Компетенция	Индикаторы достижения компетенции	Планируемые результаты обучения по дисциплине
Универсальные компетенции		
УК-3. Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	УК-3.1. Знает теоретические и методологические основы организационного руководства и выработки командной стратегии для достижения поставленной цели	Знает основы командной работы при обработке данных в задачах машинного обучения.
	УК-3.2. Умеет организовывать работу коллектива (команды) для достижения поставленной цели	Умеет организовывать работу коллектива при обработке данных в задачах машинного обучения.
	УК-3.3. Владеет конкретными методами и приемами социального взаимодействия и работы в команде, а также методами организации коллектива (команды) и его (ее) руководством	Владеет конкретными методами и приемами социального взаимодействия и работы в команде при обработке данных в задачах машинного обучения.
Общепрофессиональные компетенции		
-	-	-
Профессиональные компетенции		

Согласована на портале № 74823

ПК-1. Способен анализировать и применять методы искусственного интеллекта и машинного обучения для защиты киберфизических систем;	ПК-1.1. Нает методы искусственного интеллекта и машинного обучения для защиты киберфизических систем	Знает основные инструменты и методы анализа особенностей данных, понимает значимость таковых особенностей при обучении моделей машинного обучения. Знает основные методы для исправления недостатков в данных.
	ПК-1.2. Умеет использовать методы искусственного интеллекта и машинного обучения для защиты киберфизических систем	Умеет осуществлять построение моделей машинного обучения для решения стандартных задач информационного обеспечения и управления и применять инструменты преобразования данных для улучшения качества моделей
	ПК-1.3. Владеет методами искусственного интеллекта и машинного обучения для защиты киберфизических систем	Способен использовать средства языка python и внешних библиотек для разработки прикладных систем машинного обучения, в том числе для киберфизических систем.

4. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 4 зачетных единиц, 144 академических часов.

Распределение трудоемкости дисциплины по видам учебной деятельности представлено в таблице 4.1.

Таблица 4.1 – Трудоемкость дисциплины по видам учебной деятельности

Виды учебной деятельности	Всего часов	Семестры
		1 семестр
Контактная аудиторная работа обучающихся с преподавателем, всего	28	28
Лекционные занятия	8	8
Лабораторные занятия	20	20
Самостоятельная работа обучающихся, в т.ч. контактная внеаудиторная работа обучающихся с преподавателем, всего	116	116
Подготовка к зачету	20	20
Подготовка к тестированию	20	20
Подготовка к защите отчета по лабораторной работе	16	16
Проведение информационного поиска	44	44
Подготовка к лабораторной работе, написание отчета	16	16
Общая трудоемкость (в часах)	144	144
Общая трудоемкость (в з.е.)	4	4

5. Структура и содержание дисциплины

5.1. Разделы (темы) дисциплины и виды учебной деятельности

Структура дисциплины по разделам (темам) и видам учебной деятельности приведена в таблице 5.1.

Таблица 5.1 – Разделы (темы) дисциплины и виды учебной деятельности

Названия разделов (тем) дисциплины	Лек. зан., ч	Лаб. раб.	Сам. раб., ч	Всего часов (без экзамена)	Формируемые компетенции
1 семестр					

Согласована на портале № 74823

1 Наборы данных в задачах машинного обучения	2	8	36	46	ПК-1, УК-3
2 Несбалансированные данные	2	4	28	34	ПК-1, УК-3
3 Пропуски и аномалии данных	2	4	26	32	ПК-1, УК-3
4 Отбор и извлечение признаков	2	4	26	32	ПК-1, УК-3
Итого за семестр	8	20	116	144	
Итого	8	20	116	144	

5.2. Содержание разделов (тем) дисциплины

Содержание разделов (тем) дисциплины (в т.ч. по лекциям) приведено в таблице 5.2.

Таблица 5.2 – Содержание разделов (тем) дисциплины (в т.ч. по лекциям)

Названия разделов (тем) дисциплины	Содержание разделов (тем) дисциплины (в т.ч. по лекциям)	Трудоемкость (лекционные занятия), ч	Формируемые компетенции
1 семестр			
1 Наборы данных в задачах машинного обучения	Различие наборов данных в зависимости от задач. Характеристики и особенности данных. Специфика данных в зависимости от предметной области. Источники данных. Репозитории. Юридические аспекты сбора, использования и публикации данных. Форматы наборов данных. Метаданные. Типы данных. Смешанные данные. Влияние особенностей данных на результат. Визуализация данных. Трансформация данных.	2	ПК-1, УК-3
	Итого	2	
2 Несбалансированные данные	Проблема несбалансированности данных. Способы оценки величины несбалансированности. Методы преодоления проблемы дисбаланса. Алгоритмы удаления и дополнения данных.	2	ПК-1, УК-3
	Итого	2	
3 Пропуски и аномалии данных	Проблема наличия пропусков в данных. Способы исправления проблемы пропусков. Проблема наличия выбросов в данных. Способы нахождения и устранения выбросов.	2	ПК-1, УК-3
	Итого	2	
4 Отбор и извлечение признаков	Цели извлечения и отбора признаков. Отличие отбора от извлечения признаков. Основные методы отбора признаков. Алгоритмы фильтрации признаков. Встроенные алгоритмы отбора признаков. Алгоритмы-обертки. Алгоритмы извлечения признаков.	2	ПК-1, УК-3
	Итого	2	

Согласована на портале № 74823

Итого за семестр	8	
Итого	8	

5.3. Практические занятия (семинары)

Не предусмотрено учебным планом

5.4. Лабораторные занятия

Наименование лабораторных работ приведено в таблице 5.4.

Таблица 5.4 – Наименование лабораторных работ

Названия разделов (тем) дисциплины	Наименование лабораторных работ	Трудоемкость, ч	Формируемые компетенции
1 семестр			
1 Наборы данных в задачах машинного обучения	Разведочный анализ данных	4	ПК-1, УК-3
	Алгоритмы трансформации данных	4	ПК-1, УК-3
	Итого	8	
2 Несбалансированные данные	Алгоритмы преодоления проблемы дисбаланса данных	4	ПК-1, УК-3
	Итого	4	
3 Пропуски и аномалии данных	Методы устранения пропусков и преодоления аномалий	4	ПК-1, УК-3
	Итого	4	
4 Отбор и извлечение признаков	Алгоритмы отбора признаков	4	ПК-1, УК-3
	Итого	4	
Итого за семестр		20	
Итого		20	

5.5. Курсовой проект / курсовая работа

Не предусмотрено учебным планом

5.6. Самостоятельная работа

Виды самостоятельной работы, трудоемкость и формируемые компетенции представлены в таблице 5.6.

Таблица 5.6. – Виды самостоятельной работы, трудоемкость и формируемые компетенции

Названия разделов (тем) дисциплины	Виды самостоятельной работы	Трудоемкость, ч	Формируемые компетенции	Формы контроля
1 семестр				

Согласована на портале № 74823

1 Наборы данных в задачах машинного обучения	Подготовка к зачету	8	ПК-1, УК-3	Зачёт
	Подготовка к тестированию	8	ПК-1, УК-3	Тестирование
	Подготовка к защите отчета по лабораторной работе	4	ПК-1, УК-3	Защита отчета по лабораторной работе
	Проведение информационного поиска	12	ПК-1, УК-3	Информационный поиск
	Подготовка к лабораторной работе, написание отчета	4	ПК-1, УК-3	Лабораторная работа
	Итого	36		
2 Несбалансированные данные	Подготовка к зачету	4	ПК-1, УК-3	Зачёт
	Подготовка к тестированию	4	ПК-1, УК-3	Тестирование
	Подготовка к защите отчета по лабораторной работе	4	ПК-1, УК-3	Защита отчета по лабораторной работе
	Проведение информационного поиска	12	ПК-1, УК-3	Информационный поиск
	Подготовка к лабораторной работе, написание отчета	4	ПК-1, УК-3	Лабораторная работа
	Итого	28		
3 Пропуски и аномалии данных	Подготовка к зачету	4	ПК-1, УК-3	Зачёт
	Подготовка к тестированию	4	ПК-1, УК-3	Тестирование
	Подготовка к защите отчета по лабораторной работе	4	ПК-1, УК-3	Защита отчета по лабораторной работе
	Проведение информационного поиска	10	ПК-1, УК-3	Информационный поиск
	Подготовка к лабораторной работе, написание отчета	4	ПК-1, УК-3	Лабораторная работа
	Итого	26		

Согласована на портале № 74823

4 Отбор и извлечение признаков	Подготовка к зачету	4	ПК-1, УК-3	Зачёт
	Подготовка к тестированию	4	ПК-1, УК-3	Тестирование
	Подготовка к защите отчета по лабораторной работе	4	ПК-1, УК-3	Защита отчета по лабораторной работе
	Проведение информационного поиска	10	ПК-1, УК-3	Информационный поиск
	Подготовка к лабораторной работе, написание отчета	4	ПК-1, УК-3	Лабораторная работа
	Итого	26		
Итого за семестр		116		
Итого		116		

5.7. Соответствие компетенций, формируемых при изучении дисциплины, и видов учебной деятельности

Соответствие компетенций, формируемых при изучении дисциплины, и видов учебной деятельности представлено в таблице 5.7.

Таблица 5.7 – Соответствие компетенций, формируемых при изучении дисциплины, и видов занятий

Формируемые компетенции	Виды учебной деятельности			Формы контроля
	Лек. зан.	Лаб. раб.	Сам. раб.	
ПК-1	+	+	+	Зачёт, Защита отчета по лабораторной работе, Информационный поиск, Лабораторная работа, Тестирование
УК-3	+	+	+	Зачёт, Защита отчета по лабораторной работе, Информационный поиск, Лабораторная работа, Тестирование

6. Рейтинговая система для оценки успеваемости обучающихся

6.1. Балльные оценки для форм контроля

Балльные оценки для форм контроля представлены в таблице 6.1.

Таблица 6.1 – Балльные оценки

Формы контроля	Максимальный балл на 1-ую КТ с начала семестра	Максимальный балл за период между 1КТ и 2КТ	Максимальный балл за период между 2КТ и на конец семестра	Всего за семестр
1 семестр				
Зачёт	0	0	30	30
Защита отчета по лабораторной работе	5	5	5	15
Лабораторная работа	5	10	10	25
Тестирование	0	0	15	15
Информационный поиск	5	5	5	15
Итого максимум за период	15	20	65	100

Согласована на портале № 74823

Нарастающим итогом	15	35	100	100

6.2. Пересчет баллов в оценки за текущий контроль

Пересчет баллов в оценки за текущий контроль представлен в таблице 6.2.

Таблица 6.2 – Пересчет баллов в оценки за текущий контроль

Баллы на дату текущего контроля	Оценка
≥ 90% от максимальной суммы баллов на дату ТК	5
От 70% до 89% от максимальной суммы баллов на дату ТК	4
От 60% до 69% от максимальной суммы баллов на дату ТК	3
< 60% от максимальной суммы баллов на дату ТК	2

6.3. Пересчет суммы баллов в традиционную и международную оценку

Пересчет суммы баллов в традиционную и международную оценку представлен в таблице 6.3.

Таблица 6.3 – Пересчет суммы баллов в традиционную и международную оценку

Оценка	Итоговая сумма баллов, учитывает успешно сданный экзамен	Оценка (ECTS)
5 (отлично) (зачтено)	90 – 100	A (отлично)
4 (хорошо) (зачтено)	85 – 89	B (очень хорошо)
	75 – 84	C (хорошо)
	70 – 74	D (удовлетворительно)
3 (удовлетворительно) (зачтено)	65 – 69	
	60 – 64	E (посредственно)
2 (неудовлетворительно) (не зачтено)	Ниже 60 баллов	F (неудовлетворительно)

7. Учебно-методическое и информационное обеспечение дисциплины

7.1. Основная литература

1. Воронов, М. В. Системы искусственного интеллекта : учебник и практикум для вузов / М. В. Воронов, В. И. Пименов, И. А. Небаев. — Москва : Издательство Юрайт, 2022. — 256 с. [Электронный ресурс]: — Режим доступа: https://urait.ru/bcode/485440.

7.2. Дополнительная литература

1. Бурков, А. Инженерия машинного обучения / А. Бурков ; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2022. — 306 с. [Электронный ресурс]: — Режим доступа: https://e.lanbook.com/book/314834.

2. Монарх, Р. Машинное обучение с участием человека / Р. Монарх ; перевод с английского В. И. Бахура. — Москва : ДМК Пресс, 2022. — 498 с. [Электронный ресурс]: — Режим доступа: https://e.lanbook.com/book/241211.

3. Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль ; перевод с английского А. А. Слинкина. — 2-е изд. — Москва : ДМК Пресс, 2018. — 652 с. [Электронный ресурс]: — Режим доступа: https://e.lanbook.com/book/107901.

7.3. Учебно-методические пособия

7.3.1. Обязательные учебно-методические пособия

1. Демидова, Л. А. Разведочный анализ данных. Python : учебно-методическое пособие / Л. А. Демидова. — Москва : РТУ МИРЭА, 2022 — Часть 1 — 2022. [Электронный ресурс]: — Режим доступа: https://e.lanbook.com/book/310970.

7.3.2. Учебно-методические пособия для лиц с ограниченными возможностями здоровья и инвалидов

Согласована на портале № 74823

Учебно-методические материалы для самостоятельной и аудиторной работы обучающихся из числа лиц с ограниченными возможностями здоровья и инвалидов предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации.

Для лиц с нарушениями зрения:

– в форме электронного документа;

– в печатной форме увеличенным шрифтом.

Для лиц с нарушениями слуха:

– в форме электронного документа;

– в печатной форме.

Для лиц с нарушениями опорно-двигательного аппарата:

– в форме электронного документа;

– в печатной форме.

7.4. Современные профессиональные базы данных и информационные справочные системы

1. При изучении дисциплины рекомендуется обращаться к современным базам данных, информационно-справочным и поисковым системам, к которым у ТУСУРа открыт доступ: https://lib.tusur.ru/ru/resursy/bazy-dannyh.

8. Материально-техническое и программное обеспечение дисциплины

8.1. Материально-техническое и программное обеспечение для лекционных занятий

Для проведения занятий лекционного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации используется учебная аудитория с достаточным количеством посадочных мест для учебной группы, оборудованная доской и стандартной учебной мебелью. Имеются мультимедийное оборудование и учебно-наглядные пособия, обеспечивающие тематические иллюстрации по лекционным разделам дисциплины.

8.2. Материально-техническое и программное обеспечение для лабораторных работ

Аудитория Интернет-технологий и информационно-аналитической деятельности: учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий практического типа, учебная аудитория для проведения занятий лабораторного типа; 634045, Томская область, г. Томск, ул. Красноармейская, д. 146, 402 ауд.

Описание имеющегося оборудования:

- Интерактивная доска IQBoard DVT TN100;
- Проектор Optoma EH400;
- Веб-камера Logitech C920s;
- Акустическая система Yamaha;
- Комплект беспроводных микрофонов Clevermic;
- Магнитно-маркерная доска;
- Комплект специализированной учебной мебели;
- Рабочее место преподавателя.

Программное обеспечение:

- Microsoft Windows 10;
- VirtualBox;

8.3. Материально-техническое и программное обеспечение для самостоятельной работы

Для самостоятельной работы используются учебные аудитории (компьютерные классы), расположенные по адресам:

- 634050, Томская область, г. Томск, Ленина проспект, д. 40, 233 ауд.;
- 634045, Томская область, г. Томск, ул. Красноармейская, д. 146, 209 ауд.;
- 634034, Томская область, г. Томск, Вершинина улица, д. 47, 126 ауд.;
- 634034, Томская область, г. Томск, Вершинина улица, д. 74, 207 ауд.

Описание имеющегося оборудования:

- учебная мебель;
- компьютеры;

Согласована на портале № 74823

- компьютеры подключены к сети «Интернет» и обеспечивают доступ в электронную информационно-образовательную среду ТУСУРа.

Перечень программного обеспечения:

- Microsoft Windows;
- OpenOffice;
- Kaspersky Endpoint Security 10 для Windows;
- 7-Zip;
- Google Chrome.

8.4. Материально-техническое обеспечение дисциплины для лиц с ограниченными возможностями здоровья и инвалидов

Освоение дисциплины лицами с ограниченными возможностями здоровья и инвалидами осуществляется с использованием средств обучения общего и специального назначения.

При занятиях с обучающимися с **нарушениями слуха** предусмотрено использование звукоусиливающей аппаратуры, мультимедийных средств и других технических средств приема/передачи учебной информации в доступных формах, мобильной системы преподавания для обучающихся с инвалидностью, портативной индукционной системы. Учебная аудитория, в которой занимаются обучающиеся с нарушением слуха, оборудована компьютерной техникой, аудиотехникой, видеотехникой, электронной доской, мультимедийной системой.

При занятиях с обучающимися с **нарушениями зрения** предусмотрено использование в лекционных и учебных аудиториях возможности просмотра удаленных объектов (например, текста на доске или слайда на экране) при помощи видеоувеличителей для комфортного просмотра.

При занятиях с обучающимися с **нарушениями опорно-двигательного аппарата** используются альтернативные устройства ввода информации и другие технические средства приема/передачи учебной информации в доступных формах, мобильной системы обучения для людей с инвалидностью.

9. Оценочные материалы и методические рекомендации по организации изучения дисциплины

9.1. Содержание оценочных материалов для текущего контроля и промежуточной аттестации

Для оценки степени сформированности и уровня освоения закрепленных за дисциплиной компетенций используются оценочные материалы, представленные в таблице 9.1.

Таблица 9.1 – Формы контроля и оценочные материалы

Названия разделов (тем) дисциплины	Формируемые компетенции	Формы контроля	Оценочные материалы (ОМ)
1 Наборы данных в задачах машинного обучения	ПК-1, УК-3	Зачёт	Перечень вопросов для зачета
		Защита отчета по лабораторной работе	Примерный перечень вопросов для защиты лабораторных работ
		Лабораторная работа	Темы лабораторных работ
		Тестирование	Примерный перечень тестовых заданий
		Информационный поиск	Вид информационного поиска, перечень задач информационного поиска и содержание задания по видам поиска

Согласована на портале № 74823

2 Несбалансированные данные	ПК-1, УК-3	Зачёт	Перечень вопросов для зачета
		Защита отчета по лабораторной работе	Примерный перечень вопросов для защиты лабораторных работ
		Лабораторная работа	Темы лабораторных работ
		Тестирование	Примерный перечень тестовых заданий
		Информационный поиск	Вид информационного поиска, перечень задач информационного поиска и содержание задания по видам поиска
3 Пропуски и аномалии данных	ПК-1, УК-3	Зачёт	Перечень вопросов для зачета
		Защита отчета по лабораторной работе	Примерный перечень вопросов для защиты лабораторных работ
		Лабораторная работа	Темы лабораторных работ
		Тестирование	Примерный перечень тестовых заданий
		Информационный поиск	Вид информационного поиска, перечень задач информационного поиска и содержание задания по видам поиска
4 Отбор и извлечение признаков	ПК-1, УК-3	Зачёт	Перечень вопросов для зачета
		Защита отчета по лабораторной работе	Примерный перечень вопросов для защиты лабораторных работ
		Лабораторная работа	Темы лабораторных работ
		Тестирование	Примерный перечень тестовых заданий
		Информационный поиск	Вид информационного поиска, перечень задач информационного поиска и содержание задания по видам поиска

Шкала оценки сформированности отдельных планируемых результатов обучения по дисциплине приведена в таблице 9.2.

Таблица 9.2 – Шкала оценки сформированности планируемых результатов обучения по дисциплине

Оценка	Баллы за ОМ	Формулировка требований к степени сформированности планируемых результатов обучения		
		знать	уметь	владеть

Согласована на портале № 74823

2 (неудовлетворительно)	< 60% от максимальной суммы баллов	отсутствие знаний или фрагментарные знания	отсутствие умений или частично освоенное умение	отсутствие навыков или фрагментарные применение навыков
3 (удовлетворительно)	от 60% до 69% от максимальной суммы баллов	общие, но не структурированные знания	в целом успешно, но не систематически осуществляемое умение	в целом успешное, но не систематическое применение навыков
4 (хорошо)	от 70% до 89% от максимальной суммы баллов	сформированные, но содержащие отдельные проблемы знания	в целом успешное, но содержащие отдельные пробелы умение	в целом успешное, но содержащие отдельные пробелы применение навыков
5 (отлично)	≥ 90% от максимальной суммы баллов	сформированные систематические знания	сформированное умение	успешное и систематическое применение навыков

Шкала комплексной оценки сформированности компетенций приведена в таблице 9.3.

Таблица 9.3 – Шкала комплексной оценки сформированности компетенций

Оценка	Формулировка требований к степени компетенции
2 (неудовлетворительно)	Не имеет необходимых представлений о проверяемом материале или Знать на уровне **ориентирования**, представлений. Обучающийся знает основные признаки или термины изучаемого элемента содержания, их отнесенность к определенной науке, отрасли или объектам, узнает в текстах, изображениях или схемах и знает, к каким источникам нужно обращаться для более детального его усвоения.
3 (удовлетворительно)	Знать и уметь на **репродуктивном** уровне. Обучающихся знает изученный элемент содержания репродуктивно: произвольно воспроизводит свои знания устно, письменно или в демонстрируемых действиях.
4 (хорошо)	Знать, уметь, владеть на **аналитическом** уровне. Зная на репродуктивном уровне, указывать на особенности и взаимосвязи изученных объектов, на их достоинства, ограничения, историю и перспективы развития и особенности для разных объектов усвоения.
5 (отлично)	Знать, уметь, владеть на **системном** уровне. Обучающийся знает изученный элемент содержания системно, произвольно и доказательно воспроизводит свои знания устно, письменно или в демонстрируемых действиях, учитывая и указывая связи и зависимости между этим элементом и другими элементами содержания дисциплины, его значимость в содержании дисциплины.

9.1.1. Примерный перечень тестовых заданий

1. Какие из перечисленных лицензий семейства Creative Commons не позволяют использовать данные в коммерческих целях?
    1. CC BY-NC

Согласована на портале № 74823

2. CC BY-SA
3. CC BY-ND
4. CC BY

2. Какой из представленных наборов данных может быть опубликован в открытом доступе согласно российскому законодательству?
    1. необезличенные истории болезни пациентов
    2. данные о производственном процессе продукта, отнесенные предприятием к коммерческой тайне
    3. статистика посещений покупателями торгового центра
    4. данные о расположении воинских частей
3. Какое из представленных свойств коэффициента корреляции не верно?
    1. Если коэффициента корреляции равен нулю или близок к нему, то очевидной зависимости между признаками нет.
    2. Если коэффициента корреляции больше нуля, то большим значениям первого признака, как правило, соответствуют большие значения второго признака.
    3. Если коэффициента корреляции меньше нуля, то большим значениям первого признака, как правило, соответствуют меньшие значения второго признака.
    4. Чем ближе значение модуля коэффициента корреляции к нулю, тем легче предсказать по значению одного признака значение второго.
4. Какой из методов определения выбросов связан с анализом отдельных признаков, а не всех признаков совместно
    1. Определение выбросов с помощью кластеризации
    2. Определение выбросов с помощью геометрических методов
    3. Определение выбросов с помощью метрических методов
    4. Определение выброса по критерию Шовене
5. В чем заключается назначение метода главных компонент при обработке наборов данных?
    1. поиск и удаление наименее информативных признаков
    2. выделение наиболее полезных для построения модели экземпляров данных
    3. снижение размерности данных путем создания новых признаков из исходных
    4. преобразование категориальных данных в числовой формат
6. Какое из действий не позволяет решить проблему пропусков значений признаков в наборе данных?
    1. Исключение данных с пропусками
    2. Внесение недостающих значений
    3. Замена недостающих значений
    4. Игнорирование пропусков
7. Как называется метод, направленный на устранение дисбаланса по классам, с помощью случайного исключения экземпляров класса большинства?
    1. Random over-sampling
    2. Down-sampling
    3. Random under-sampling
    4. Up-sampling
8. Какие типы признаков могут присутствовать в наборе данных?
    1. количественные, бинарные, порядковые и категориальные
    2. количественные, бинарные, порядковые и интегральные
    3. количественные, бинарные, интегральные и категориальные
    4. целочисленные, бинарные, интегральные и категориальные
9. Какие из типов метрик НЕ ОТНОСЯТСЯ к метрикам качества классификации?
    1. Метрики, направленные на минимизацию числа ошибок (например, F1-мера)
    2. Метрики, основанные на оптимизации конверсии (например, RPV)
    3. Метрики, основанные на вероятностном понимании ошибки (например, Loss)
    4. Метрики, основанные на оценивании разделимости экземпляров (например, AUC)
10. С какой целью используются алгоритм SMOTE?
    1. устранение пропусков путем заполнения усредненными значениями
    2. генерация дополнительных данных на основе существующих

Согласована на портале № 74823

3. отбор признаков в режиме фильтрации
4. сокращение числа экземпляров данных путем устранения соседей

11. Чему будет равен коэффициент дисбаланса для набора данных, в котором первый класс насчитывает 168 экземпляров, второй – 56 экземпляров, третий – 2452?
    1. 43,79
    2. 0,02
    3. 0,09
    4. 10,95
12. Какая команда библиотеки pandas применяются для поиска значений «NaN»?
    1. findnull()
    2. find(nan)
    3. info(nan)
    4. isnull()
13. Какое из действий можно отнести к отличительной черте алгоритмов, реализующих отбор признаков в режиме "обертки"?
    1. выбор признаков, имеющих наибольшую корреляцию с выходной переменной
    2. оценка подмножества выбранных признаков на основе меры качества построенной модели
    3. уменьшение размерности набора путем формирования случайных пар признаков и создания одного нового признака из каждой пары
    4. циклический расчет взаимной информации между подмножеством признаков и выходной переменной
14. Какая команда Python может использоваться для вывода значений всех столбцов первых строк датафрейма Pandas?
    1. df.head
    2. print(df)
    3. df.info
    4. print(df, 5)
15. Какая команда Python позволяет обратиться к столбцу “name_column” в датафрейме Pandas df?
    1. df[name_column]
    2. df[‘name_column’]
    3. name_column.df
    4. name_column[df]
16. Как называется пара экземпляров противоположных классов, являющиеся ближайшими соседями, при использовании в одном из методов недосемплирования?
    1. связи каскадной балансировки
    2. связи Томека
    3. корреляционная связь
    4. экземпляры миноритарного класса
17. Для какого метода повышения дискретизации, количество генерируемых экземпляров для рассматриваемого экземпляра пропорционально его сложности?
    1. SMOTE
    2. KM-SMOTE
    3. ADASYN
    4. EasyEnsemble
18. Для чего предназначен коэффициент k, используемый в участке кода?

```
iqr=q3-q1
k=1.5
upper_bound=q3+(k*iqr)
lower_bound=q1-(k*iqr)
```

    1. коэффициент, определяющий величину верхней и нижней границы определения выбросов
    2. коэффициент выбора количества исключаемых значений
    3. коэффициент подбора квартилей
    4. коэффициент равный верхнему и нижнему пределу значений фильтрации выборки

Согласована на портале № 74823

19. Почему несбалансированность данных в наборе является проблемой при решении задач машинного обучения?
    1. экземпляры целевого класса часто пересекаются с областями классов меньшинства
    2. большой размер выборки и слабая разделимость
    3. процесс обучения показывает высокую общую точность, однако классифицирует экземпляры класса меньшинств
    4. модель показывает высокую общую точность, однако некорректно классифицирует экземпляры меньшего класса
20. Для решения какой задачи используется нормализация данных?
    1. приведение данных к единому масштабу
    2. устранение пропусков (значений NaN)
    3. извлечение из смешанных данных числовых значений
    4. приведение данных к формату, воспринимаемому решающим алгоритмом

9.1.2. Перечень вопросов для зачета

1. Методология ведения проектов интеллектуального анализа данных CRISP-DM.
2. Источники данных. Репозитории
3. Типы лицензий на наборы данных.
4. Типы данных
5. Проблема несбалансированности данных. Способы оценки величины несбалансированности.
6. Методы преодоления проблемы дисбаланса.
7. Преодоление проблемы дисбаланса: алгоритмы удаления и дополнения данных.
8. Проблема наличия пропусков в данных. Типы пропусков в данных.
9. Способы исправления проблемы пропусков
10. Проблема наличия выбросов в данных. Способы нахождения и устранения выбросов
11. Основные методы отбора признаков
12. Алгоритмы фильтрации признаков
13. Встроенные алгоритмы отбора признаков
14. Алгоритмы-обертки
15. “Жадные” алгоритмы отбора признаков
16. Генетический алгоритм отбора признаков
17. Алгоритмы извлечения (генерации) признаков.
18. Метод главных компонент (PCA).

9.1.3. Примерный перечень вопросов для защиты лабораторных работ

1. В чем заключается назначение метода главных компонент при обработке наборов данных?
2. Какие действия позволяют решить проблему пропусков значений признаков в наборе данных?
3. Как называется метод, направленный на устранение дисбаланса по классам, с помощью случайного исключения экземпляров класса большинства?
4. Какие типы признаков могут присутствовать в наборе данных?
5. Какие из типов метрик НЕ ОТНОСЯТСЯ к метрикам качества классификации?
6. С какой целью используются алгоритм SMOTE?
7. Чему будет равен коэффициент дисбаланса для набора данных, в котором первый класс насчитывает 168 экземпляров, второй – 56 экземпляров, третий – 2452?
8. Какая команда библиотеки pandas применяются для поиска значений «NaN»?
9. Какое из действий можно отнести к отличительной черте алгоритмов, реализующих отбор признаков в режиме "обертки"?
10. Какая команда Python может использоваться для вывода значений всех столбцов первых строк датафрейма Pandas?
11. Какая команда Python позволяет обратиться к столбцу “name_column” в датафрейме Pandas df?
12. Как называется пара экземпляров противоположных классов, являющиеся ближайшими соседями, при использовании в одном из методов недосемплирования?

Согласована на портале № 74823

13. Для какого метода повышения дискретизации, количество генерируемых экземпляров для рассматриваемого экземпляра пропорционально его сложности?
14. Для чего предназначен коэффициент k, используемый в участке кода?
    iqr=q3-q1
    k=1.5
    upper_bound=q3+(k*iqr)
    lower_bound=q1-(k*iqr)
15. Почему несбалансированность данных в наборе является проблемой при решении задач машинного обучения?
16. Для решения какой задачи используется нормализация данных?

9.1.4. Темы лабораторных работ

1. Разведочный анализ данных
2. Алгоритмы трансформации данных
3. Алгоритмы преодоления проблемы дисбаланса данных
4. Методы устранения пропусков и преодоления аномалий
5. Алгоритмы отбора признаков

9.1.5. Вид информационного поиска, перечень задач информационного поиска и содержание задания по видам поиска

При выполнении задач информационного поиска производится сбор, формирование данных, анализ литературных источников по проблемам, заявленным в задачах информационного поиска. Информационный поиск требуется для расширения и закрепления знаний, сформированных в рамках лекционных занятий, определения актуальных тенденций и современных подходов в рассматриваемых предметных областях.

1. Методология ведения проектов интеллектуального анализа данных CRISP-DM.
2. Источники данных. Репозитории
3. Типы лицензий на наборы данных.
4. Типы данных
5. Методы преодоления проблемы дисбаланса.
6. Преодоление проблемы дисбаланса: алгоритмы удаления и дополнения данных.
7. Способы исправления проблемы пропусков
8. Способы нахождения и устранения выбросов в данных
9. Алгоритмы фильтрации признаков
10. Встроенные алгоритмы отбора признаков
11. Алгоритмы-обертки
12. “Жадные” алгоритмы отбора признаков
13. Генетический алгоритм отбора признаков
14. Алгоритмы извлечения (генерации) признаков.
15. Метод главных компонент (PCA).

9.2. Методические рекомендации

Учебный материал излагается в форме, предполагающей самостоятельное мышление студентов, самообразование. При этом самостоятельная работа студентов играет решающую роль в ходе всего учебного процесса.

Начать изучение дисциплины необходимо со знакомства с рабочей программой, списком учебно-методического и программного обеспечения. Самостоятельная работа студента включает работу с учебными материалами, выполнение контрольных мероприятий, предусмотренных учебным планом.

В процессе изучения дисциплины для лучшего освоения материала необходимо регулярно обращаться к рекомендуемой литературе и источникам, указанным в учебных материалах; пользоваться через кабинет студента на сайте Университета образовательными ресурсами электронно-библиотечной системы, а также общедоступными интернет-порталами, содержащими научно-популярные и специализированные материалы, посвященные различным аспектам учебной дисциплины.

При самостоятельном изучении тем следуйте рекомендациям:

Согласована на портале № 74823

– чтение или просмотр материала осуществляйте со скоростью, достаточной для индивидуального понимания и освоения материала, выделяя основные идеи; на основании изученного составить тезисы. Освоив материал, попытаться соотнести теорию с примерами из практики;

– если в тексте встречаются незнакомые или малознакомые термины, следует выяснить их значение для понимания дальнейшего материала;

– осмысливайте прочитанное и изученное, отвечайте на предложенные вопросы.

Студенты могут получать индивидуальные консультации, в т.ч. с использованием средств телекоммуникации.

По дисциплине могут проводиться дополнительные занятия, в т.ч. в форме вебинаров. Расписание вебинаров и записи вебинаров публикуются в электронном курсе / электронном журнале по дисциплине.

9.3. Требования к оценочным материалам для лиц с ограниченными возможностями здоровья и инвалидов

Для лиц с ограниченными возможностями здоровья и инвалидов предусмотрены дополнительные оценочные материалы, перечень которых указан в таблице 9.4.

Таблица 9.4 – Дополнительные материалы оценивания для лиц с ограниченными возможностями здоровья и инвалидов

Категории обучающихся	Виды дополнительных оценочных материалов	Формы контроля и оценки результатов обучения
С нарушениями слуха	Тесты, письменные самостоятельные работы, вопросы к зачету, контрольные работы	Преимущественно письменная проверка
С нарушениями зрения	Собеседование по вопросам к зачету, опрос по терминам	Преимущественно устная проверка (индивидуально)
С нарушениями опорно-двигательного аппарата	Решение дистанционных тестов, контрольные работы, письменные самостоятельные работы, вопросы к зачету	Преимущественно дистанционными методами
С ограничениями по общемедицинским показаниям	Тесты, письменные самостоятельные работы, вопросы к зачету, контрольные работы, устные ответы	Преимущественно проверка методами, определяющимися исходя из состояния обучающегося на момент проверки

9.4. Методические рекомендации по оценочным материалам для лиц с ограниченными возможностями здоровья и инвалидов

Для лиц с ограниченными возможностями здоровья и инвалидов предусматривается доступная форма предоставления заданий оценочных средств, а именно:

– в печатной форме;

– в печатной форме с увеличенным шрифтом;

– в форме электронного документа;

– методом чтения ассистентом задания вслух;

– предоставление задания с использованием сурдоперевода.

Лицам с ограниченными возможностями здоровья и инвалидам увеличивается время на подготовку ответов на контрольные вопросы. Для таких обучающихся предусматривается доступная форма предоставления ответов на задания, а именно:

– письменно на бумаге;

– набор ответов на компьютере;

– набор ответов с использованием услуг ассистента;

– представление ответов устно.

Процедура оценивания результатов обучения лиц с ограниченными возможностями здоровья и инвалидов по дисциплине предусматривает предоставление информации в формах,

Согласована на портале № 74823

адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

– в форме электронного документа;

– в печатной форме увеличенным шрифтом.

Для лиц с нарушениями слуха:

– в форме электронного документа;

– в печатной форме.

Для лиц с нарушениями опорно-двигательного аппарата:

– в форме электронного документа;

– в печатной форме.

При необходимости для лиц с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения может проводиться в несколько этапов.

Согласована на портале № 74823

ЛИСТ СОГЛАСОВАНИЯ

Рассмотрена и одобрена на заседании кафедры КИБЭВС
протокол № 1 от «24» 1 2023 г.

СОГЛАСОВАНО:

Должность	Инициалы, фамилия	Подпись
Заведующий выпускающей каф. КИБЭВС	А.А. Шелупанов	Согласовано, c53e145e-8b20-45aa-9347-a5e4dbb90e8d
Заведующий обеспечивающей каф. КИБЭВС	А.А. Шелупанов	Согласовано, c53e145e-8b20-45aa-9347-a5e4dbb90e8d
И.О. начальника учебного управления	И.А. Лариошина	Согласовано, c3195437-a02f-4972-a7c6-ab6ee1f21e73

ЭКСПЕРТЫ:

Доцент, каф. КИБЭВС	А.А. Конев	Согласовано, 81687a04-85ce-4835-9e1e-9934a6085fdd
Доцент, каф. КИБЭВС	А.Ю. Якимук	Согласовано, 4ffdf265-fb78-4863-b293-f03438cb07cc

РАЗРАБОТАНО:

Доцент, каф. БИС	И.А. Рахманенко	Разработано, 438e5305-e83a-40ae-b333-7c84f2fc4661

Согласована на портале № 74823