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 учебной работе
_____П.В. Сенченко
«23» _____12_____ 2020 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

НАБОРЫ ДАННЫХ В ЗАДАЧАХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Уровень образования: **высшее образование - специалитет**
Направление подготовки / специальность: **10.05.04 Информационно-аналитические системы безопасности**
Направленность (профиль) / специализация: **Информационная безопасность финансовых и экономических структур**
Форма обучения: **очная**
Факультет: **Факультет безопасности (ФБ)**
Кафедра: **Кафедра безопасности информационных систем (БИС)**
Курс: **2**
Семестр: **3**
Учебный план набора 2021 года

Объем дисциплины и виды учебной деятельности

Виды учебной деятельности	3 семестр	Всего	Единицы
Лекционные занятия	8	8	часов
Практические занятия	20	20	часов
Самостоятельная работа	44	44	часов
Общая трудоемкость (включая промежуточную аттестацию)	72	72	часов
	2	2	з.е.

Формы промежуточной аттестация	Семестр
Зачет	3

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Сенченко П.В.
Должность: Проректор по УР
Дата подписания: 23.12.2020
Уникальный программный ключ:
a1119608-cdff-4455-b54e-5235117c185c

Томск

Согласована на портале № 56292

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 преодолению и исправлению проблем анализируемых данных, необходимых студентам 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: Обязательная часть.
Модуль дисциплин: Модуль специальности (special hard skills - SHS).
Индекс дисциплины: Б1.О.03.09.
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		
-	-	-
Общепрофессиональные компетенции		

Согласована на портале № 56292

ОПК-15. Способен осуществлять автоматизированную информационно-аналитическую поддержку процессов принятия решений на базе ситуационных центров	ОПК-15.1. Знает технологию и алгоритмы работы ситуационных центров различной направленности, современные информационно-аналитические системы, используемые для автоматизации работы ситуационных центров	Способен проводить анализ особенностей данных, оценивать значимость таковых особенностей, прогнозировать их влияние на результат построения моделей машинного обучения. Способен выбирать соответствующие инструменты для исправления недостатков в данных.
	ОПК-15.2. Умеет использовать модели данных и знаний для решения стандартных задач информационного обеспечения и управления, применять информационно-аналитические системы в информационно-аналитической деятельности	Умеет осуществлять построение моделей машинного обучения для решения стандартных задач информационного обеспечения и управления и применять инструменты преобразования данных для улучшения качества моделей.
	ОПК-15.3. Владеет методами и средствами разработки прикладных систем поддержки баз данных и знаний	Способен использовать средства языка python и внешних библиотек для разработки прикладных систем поддержки баз данных и знаний
Профессиональные компетенции		
-	-	-

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 2 зачетных единиц, 72 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
		3 семестр
Контактная аудиторная работа обучающихся с преподавателем, всего	28	28
Лекционные занятия	8	8
Практические занятия	20	20
Самостоятельная работа обучающихся, в т.ч. контактная внеаудиторная работа обучающихся с преподавателем, всего	44	44
Подготовка к зачету	10	10
Подготовка к тестированию	8	8
Проведение информационного поиска	2	2
Подготовка к защите отчета по практическому занятию	4	4
Выполнение практического задания	14	14
Выполнение индивидуального задания	4	4
Подготовка к защите отчета по индивидуальному заданию	2	2

Согласована на портале № 56292

Общая трудоемкость (в часах)	72	72
Общая трудоемкость (в з.е.)	2	2

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	Прак. зан., ч	Сам. раб., ч	Всего часов (без экзамена)	Формируемые компетенции
3 семестр					
1 Наборы данных в задачах машинного обучения	2	6	9	17	ОПК-15
2 Несбалансированные данные	2	4	9	15	ОПК-15
3 Пропуски и аномалии данных	2	4	9	15	ОПК-15
4 Отбор и извлечение признаков	2	6	17	25	ОПК-15
Итого за семестр	8	20	44	72	
Итого	8	20	44	72	

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
3 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	ОПК-15
	Итого	2	
2 Несбалансированные данные	Проблема несбалансированности данных. Способы оценки величины несбалансированности. Методы преодоления проблемы дисбаланса. Алгоритмы удаления и дополнения данных.	2	ОПК-15
	Итого	2	

Согласована на портале № 56292

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	ОПК-15
	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	ОПК-15
	Итого	2	
Итого за семестр		8	
Итого		8	

5.3. Практические занятия (семинары)

Наименование практических занятий (семинаров) приведено в таблице 5.3.

Таблица 5.3 – Наименование практических занятий (семинаров)

Названия разделов (тем) дисциплины	Наименование практических занятий (семинаров)	Трудоемкость, ч	Формируемые компетенции
3 семестр			
1 Наборы данных в задачах машинного обучения	Знакомство с репозиториями	2	ОПК-15
	Алгоритмы трансформации данных	4	ОПК-15
	Итого	6	
2 Несбалансированные данные	Алгоритмы преодоления проблемы дисбаланса данных	4	ОПК-15
	Итого	4	
3 Пропуски и аномалии данных	Методы устранения пропусков	2	ОПК-15
	Методы преодоления проблемы выбросов	2	ОПК-15
	Итого	4	
4 Отбор и извлечение признаков	Алгоритмы отбора признаков	4	ОПК-15
	Построение модели машинного обучения на тренировочных данных	2	ОПК-15
	Итого	6	
Итого за семестр		20	
Итого		20	

5.4. Лабораторные занятия

Не предусмотрено учебным планом

5.5. Курсовой проект / курсовая работа

Не предусмотрено учебным планом

5.6. Самостоятельная работа

Виды самостоятельной работы, трудоемкость и формируемые компетенции представлены в

Согласована на портале № 56292

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
3 семестр				
1 Наборы данных в задачах машинного обучения	Подготовка к зачету	2	ОПК-15	Зачёт
	Подготовка к тестированию	2	ОПК-15	Тестирование
	Проведение информационного поиска	2	ОПК-15	Информационный поиск
	Подготовка к защите отчета по практическому занятию	1	ОПК-15	Защита отчета по практическому занятию
	Выполнение практического задания	2	ОПК-15	Практическое задание
	Итого	9		
2 Несбалансированные данные	Подготовка к зачету	2	ОПК-15	Зачёт
	Подготовка к тестированию	2	ОПК-15	Тестирование
	Выполнение практического задания	4	ОПК-15	Практическое задание
	Подготовка к защите отчета по практическому занятию	1	ОПК-15	Защита отчета по практическому занятию
	Итого	9		
3 Пропуски и аномалии данных	Подготовка к зачету	2	ОПК-15	Зачёт
	Подготовка к тестированию	2	ОПК-15	Тестирование
	Подготовка к защите отчета по практическому занятию	1	ОПК-15	Защита отчета по практическому занятию
	Выполнение практического задания	4	ОПК-15	Практическое задание
	Итого	9		

Согласована на портале № 56292

4 Отбор и извлечение признаков	Подготовка к зачету	4	ОПК-15	Зачёт
	Подготовка к тестированию	2	ОПК-15	Тестирование
	Выполнение индивидуального задания	4	ОПК-15	Индивидуальное задание
	Подготовка к защите отчета по практическому занятию	1	ОПК-15	Защита отчета по практическому занятию
	Выполнение практического задания	4	ОПК-15	Практическое задание
	Подготовка к защите отчета по индивидуальному заданию	2	ОПК-15	Защита отчета по индивидуальному заданию
	Итого	17		
Итого за семестр		44		
Итого		44		

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.	Прак. зан.	Сам. раб.	
ОПК-15	+	+	+	Зачёт, Защита отчета по практическому занятию, Защита отчета по индивидуальному заданию, Индивидуальное задание, Практическое задание, Тестирование, Информационный поиск

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
3 семестр				
Зачёт	0	0	30	30
Защита отчета по практическому занятию	2	4	0	6
Защита отчета по индивидуальному заданию	0	0	2	2

Согласована на портале № 56292

Индивидуальное задание	0	0	13	13
Практическое задание	8	16	0	24
Тестирование	5	10	5	20
Информационный поиск	5	0	0	5
Итого максимум за период	20	30	50	100
Нарастающим итогом	20	50	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. Системы искусственного интеллекта: Учебное пособие / Н. В. Замятин - 2018. 244 с. [Электронный ресурс]: — Режим доступа: https://edu.tusur.ru/publications/7269.

7.2. Дополнительная литература

1. Чио, Кларенс. Машинное обучение и безопасность [Электр.ресурс] : руководство. - М. : ДМК Пресс , 2020 [Электронный ресурс]: — Режим доступа: https://reader.lanbook.com/book/131707#1.

2. Бессмертный, Игорь Александрович. Интеллектуальные системы [Электр.ресурс] : учебник и практикум для вузов. - М. : Юрайт , 2020 [Электронный ресурс]: — Режим доступа: https://urait.ru/viewer/intellektualnye-sistemy-451101#page/1.

3. Катулев А. Н. Математические методы в системах поддержки принятия решений: Учебное пособие для вузов / А. Н. Катулев. - М. : Высшая школа, 2005. - 310[2] с (наличие в библиотеке ТУСУР - 20 экз.).

7.3. Учебно-методические пособия

7.3.1. Обязательные учебно-методические пособия

Согласована на портале № 56292

1. Базы знаний: Методические указания для выполнения практических и самостоятельных работ / И. А. Ходашинский, К. С. Сарин, А. Е. Анфилофьев - 2017. 68 с. [Электронный ресурс]: — Режим доступа: https://edu.tusur.ru/publications/7151.

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

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 практических занятий

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;
- Visio;

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.;

Согласована на портале № 56292

- 634045, Томская область, г. Томск, ул. Красноармейская, д. 146, 201 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;
- 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)

Согласована на портале № 56292

1 Наборы данных в задачах машинного обучения	ОПК-15	Зачёт	Перечень вопросов для зачета
		Защита отчета по практическому занятию	Примерный перечень вопросов для защиты практических занятий
		Практическое задание	Темы практических заданий
		Тестирование	Примерный перечень тестовых заданий
		Информационный поиск	Вид информационного поиска, перечень задач информационного поиска и содержание задания по видам поиска
2 Несбалансированные данные	ОПК-15	Зачёт	Перечень вопросов для зачета
		Защита отчета по практическому занятию	Примерный перечень вопросов для защиты практических занятий
		Практическое задание	Темы практических заданий
		Тестирование	Примерный перечень тестовых заданий
3 Пропуски и аномалии данных	ОПК-15	Зачёт	Перечень вопросов для зачета
		Защита отчета по практическому занятию	Примерный перечень вопросов для защиты практических занятий
		Практическое задание	Темы практических заданий
		Тестирование	Примерный перечень тестовых заданий
4 Отбор и извлечение признаков	ОПК-15	Зачёт	Перечень вопросов для зачета
		Защита отчета по практическому занятию	Примерный перечень вопросов для защиты практических занятий
		Защита отчета по индивидуальному заданию	Примерный перечень вопросов для защиты индивидуальных заданий
		Индивидуальное задание	Примерный перечень вариантов индивидуальных заданий
		Практическое задание	Темы практических заданий
		Тестирование	Примерный перечень тестовых заданий

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

Согласована на портале № 56292

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
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.

Согласована на портале № 56292

9.1.1. Примерный перечень тестовых заданий

1. Какие из перечисленных лицензий семейства Creative Commons не позволяют использовать данные в коммерческих целях?
   а) CC BY-NC.
   б) CC BY-SA.
   в) CC BY-NC-ND.
   г) CC BY.
2. Алгоритм SMOTE предназначен для следующей цели:
   a) устранение пропусков путем заполнения усредненными значениями;
   b) генерация дополнительных данных на основе существующих;
   c) отбор признаков в режиме фильтрации;
   d) сокращение числа экземпляров данных путем устранения соседей.
3. Основное назначение метода главных компонент заключается в следующем:
   а) поиск и удаление наименее информативных признаков;
   б) выделение наиболее полезных для построения модели экземпляров данных;
   в) снижение размерности данных путем создания новых признаков из исходных;
   г) преобразование категориальных данных в числовой формат.
4. Рассчитайте коэффициент дисбаланса набора данных, в котором первый класс насчитывает 168 экземпляров, второй – 56 экземпляров, третий – 2452.
   а) 43,79;
   б) 0,02;
   в) 0,09;
   г) 10,95
5. Какая команда библиотеки pandas применяются для поиска значений «NaN»?
   а) findnull()
   б) find(nan)
   в) info(nan)
   г) isnull()
6. Отличительной чертой алгоритмов, реализующих отбор признаков в режиме "обертки", является:
   а) выбор признаков, имеющих наибольшую корреляцию с выходной переменной;
   б) оценка подмножества выбранных признаков на основе меры качества построенной модели;
   в) уменьшение размерности набора путем формирования случайных пар признаков и создания одного нового признака из каждой пары;
   г) циклический расчет взаимной информации между подмножеством признаков и выходной переменной.
7. Верное соедините термины и их определения.
   1) Персональные данные
   2) Обезличенные данные
   3) Открытые данные
   а) Данные, которые нельзя соотнести с каким-либо конкретным субъектом
   б) Данные, опубликованные в публичном доступе под открытой лицензией, допускающей бесплатное повторное использование этих данных, их переработку и распространение;
   в) Любая информация, относящаяся к определенному или определяемому на ее основании физическому лицу
8. Нормализация данных необходима для следующей цели:
   а) приведение данных к единому масштабу;
   б) устранение пропусков (значений NaN);
   в) извлечение из смешанных данных числовых значений;
   г) приведение данных к формату, воспринимаемому решающим алгоритмом.
9. Среди представленных вариантов наборов данных найдите тот, который может быть опубликован в открытом доступе согласно российскому законодательству.
   а) необезличенные истории болезни пациентов;
   б) данные о производственном процессе продукта, отнесенные предприятием к коммерческой тайне;

Согласована на портале № 56292

в) статистика посещений покупателями торгового центра;
г) данные о расположении воинских частей.

10. Среди представленных нормативно-правовых документов выберите регулирующие публикацию данных:
а) Федеральный закон от 27.07.2006 г. №149-ФЗ «Об информации, информационных технологиях и о защите информации»;
б) Федеральный закон от 27.07.2006 N 152-ФЗ «О персональных данных»;
в) Федеральный закон от 21.11.2011 N 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации»
г) Закон РФ № 5485-1 от 21.07.1993 «О государственной тайне».

9.1.2. Перечень вопросов для зачета

1. Характеристики и особенности наборов данных.
2. Источники наборов данных.
3. Виды лицензий на наборы данных.
4. Типы данных. Способы работы с нечисловыми данными.
5. Дисбаланс данных - определение, возможные последствия, способы преодоления.
6. Принцип работы алгоритмов добавления и удаления данных. Примеры алгоритмов.
7. Пропуски в данных - определение, возможные последствия, способы нахождения и преодоления.
8. Аномалии данных - понятие, возможные последствия, способы исправления.
9. Отбор признаков - понятие, цель использования, классификация алгоритмов отбора признаков.
10. Извлечение признаков - понятие, цель использования, пример алгоритма извлечения признаков.

9.1.3. Примерный перечень вопросов для защиты практических занятий

1. Обоснуйте выбор алгоритма для выполнения задания.
2. Продемонстрируйте результаты построения модели на исходных данных.
3. Продемонстрируйте результаты построения модели на преобразованных данных.
4. Перечислите используемые библиотеки для выполнения задания.
5. Перечислите используемые команды для выполнения задания.

9.1.4. Темы практических заданий

1. Алгоритмы трансформации данных
2. Алгоритмы преодоления проблемы дисбаланса данных
3. Методы устранения пропусков
4. Методы преодоления проблемы выбросов
5. Алгоритмы отбора признаков

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

1. Зарегистрируйтесь в открытом репозитории Kaggle.
2. Ознакомьтесь с интерфейсом и функционалом репозитория.
3. Перейдите в раздел Datasets – Popular Datasets. Отрегулируйте фильтр поиска таким образом, чтобы получить в выборке только наборы данных в формате CSV.
4. Выберите набор данных для регрессии или классификации, соответствующий предметной области, указанной в вашем варианте. Данные должны иметь как минимум пять признаков, большинство из которых – числовые признаки. Обязательным условием выбора набора данных является наличие подробного описания набора данных и сформулированной задачи построения модели машинного обучения, а также статус не ниже «Bronze».
5. Представьте преподавателю отчет с описанием данных, визуализацией с помощью матрицы рассеяния и обоснованием выбора набора.

9.1.6. Примерный перечень вопросов для защиты индивидуальных заданий

Согласована на портале № 56292

1. Опишите выбранный набор данных и задачу, для решения которой он применяется.
2. Продемонстрируйте визуализацию набора данных.
3. Разъясните выбор решающего алгоритма для построения модели машинного обучения.
4. Обоснуйте выбор алгоритмов для исправления недостатков данных.
5. Приведите анализ результатов и сравнение с результатами, приведенными в открытых источниках.

9.1.7. Примерный перечень вариантов индивидуальных заданий

1. Построить модель машинного обучения (классификатор или аппроксиматор) для предметной области "Медицина".
2. Построить модель машинного обучения (классификатор или аппроксиматор) для предметной области "Транспорт".
3. Построить модель машинного обучения (классификатор или аппроксиматор) для предметной области "Информационная безопасность".
4. Построить модель машинного обучения (классификатор или аппроксиматор) для предметной области "Промышленность".
5. Построить модель машинного обучения (классификатор или аппроксиматор) для предметной области "Биология".
6. Построить модель машинного обучения (классификатор или аппроксиматор) для предметной области "Торговля".

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:

– 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 9.2.

Таблица 9.2 – Дополнительные материалы оценивания для лиц с ограниченными возможностями здоровья и инвалидов

Согласована на портале № 56292

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, 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 № 56292

ЛИСТ СОГЛАСОВАНИЯ

Рассмотрена и одобрена на заседании кафедры КИБЭВС
протокол № 11 от «14» 12 2020 г.

СОГЛАСОВАНО:

Должность	Инициалы, фамилия	Подпись
Заведующий выпускающей каф. БИС	Е.Ю. Костюченко	Согласовано, c6235dfe-234a-4234-88f9-e1597aac6463
Заведующий обеспечивающей каф. КИБЭВС	А.А. Шелупанов	Согласовано, c53e145e-8b20-45aa-9347-a5e4dbb90e8d
Начальник учебного управления	Е.В. Саврук	Согласовано, fa63922b-1fce-4a6a-845d-9ce7670b004c

ЭКСПЕРТЫ:

Доцент, каф. КИБЭВС	А.А. Конев	Согласовано, 81687a04-85ce-4835-9e1e-9934a6085fdd
Доцент, каф. КИБЭВС	К.С. Сарин	Согласовано, 68c81ca0-0954-467a-8d01-f93a0d553669

РАЗРАБОТАНО:

Старший преподаватель, каф. КИБЭВС	М.Б. Бардамова	Разработано, f30aa12e-80f6-4e6e-806e-baabff923c2a

Согласована на портале № 56292