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

АНАЛИЗ БОЛЬШИХ ДАННЫХ

Уровень образования: **высшее образование - магистратура**
Направление подготовки / специальность: **09.04.01 Информатика и вычислительная техника**
Направленность (профиль) / специализация: **Автоматизированные системы обработки информации и управления в экономике**
Форма обучения: **очная**
Факультет: **Факультет систем управления (ФСУ)**
Кафедра: **Кафедра автоматизированных систем управления (АСУ)**
Курс: **1**
Семестр: **1**
Учебный план набора 2021 года

Объем дисциплины и виды учебной деятельности

Виды учебной деятельности	1 семестр	Всего	Единицы
Лекционные занятия	36	36	часов
Лабораторные занятия	36	36	часов
Самостоятельная работа	144	144	часов
Общая трудоемкость (включая промежуточную аттестацию)	216	216	часов
	6	6	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: 23.12.2020
Уникальный программный ключ:
a1119608-cdff-4455-b54e-5235117c185c

Томск

Согласована на портале № 58037

1. Общие положения

1.1. Цели дисциплины

1. Освоение основных концепций и методов аналитики данных, особенностей областей применения и использования их как готового инструмента принятия решений при работе со структурированными и неструктурированными данными больших объемов.
2. Формирование у студентов теоретических знаний и практических навыков по вопросам анализа данных; поиска управленческих решений; освоение студентами современных математических методов машинного обучения.

1.2. Задачи дисциплины

1. Изучение основных понятий, процесса и технологий «больших данных».
2. Изучение методов статистического анализа данных.
3. Изучение методов машинного обучения «с учителем» и «без учителя».
4. Формирование у студентов знаний и умений, необходимых для эффективного управления техническими, организационными и экономическими системами.

2. Место дисциплины в структуре ОПОП

Блок дисциплин: Б1. Дисциплины (модули).
Часть блока дисциплин: Часть, формируемая участниками образовательных отношений.
Модуль дисциплин: Модуль направленности (профиля) (major).
Индекс дисциплины: Б1.В.01.05.
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 № 58037

ОПК-2. Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	ОПК-2.1. Знает современные информационно-коммуникационные и интеллектуальные технологии, инструментальные среды, программно-технические платформы для решения профессиональных задач	Знает современные программный средства для обработки и анализа данных.
	ОПК-2.2. Умеет обосновывать выбор современных информационно-коммуникационных и интеллектуальных технологий, разрабатывать оригинальные алгоритмы и программные средства для решения профессиональных задач	Умеет осуществлять выбор программных средств для обработки и анализа данных.
	ОПК-2.3. Владеет методами разработки оригинальных программных средств с использованием современных информационно-коммуникационных, в том числе и интеллектуальных технологий, для решения профессиональных задач	Имеет опыт разработки программных средств для обработки данных с использованием электронных таблиц.
Профессиональные компетенции		

ПКС-3. Способен разрабатывать варианты управленческих решений и обосновывать их выбор на основе критериев социально-экономической эффективности	ПКС-3.1. Знает варианты управленческих решений с использованием интеллектуальных методов и информационных технологий (Python, Excel (Calc , RStudio)	Знает основные принципы и технологии аналитики больших данных.
	ПКС-3.2. Умеет применять методы машинного обучения для обоснования управленческих решений (используя Python, Excel (Calc), RStudio)	Умет выполнять ранжирование для оценки многокритериальных альтернатив; составлять модели множественной линейной регрессии; осуществлять кластерный анализ методом k-средних, K-медиан; выделять сообщества в сетевых графах; строить деревья решений, ансамбли моделей.
	ПКС-3.3. Владеет методами разработки предметно-ориентированных систем поддержки принятия решений	Имеет опыт применения методов машинного обучения «с учителем» и «без учителя» для обоснования выбора вариантов управленческих решений.
ПКС-5. Способен собирать, обрабатывать и интерпретировать данные современных научных исследований	ПКС-5.1. Знает данные научных исследований на основе интеллектуальных методов и информационных технологий (Colab, Python, библиотеки Keras, Numpy, фреймворк TensorFlow , система Clips , Excel (Calc), RStudio).	Знает теоретического основы статистического анализа данных, машинного обучения.
	ПКС-5.2. Умеет применять методы и инструменты нейронных сетей для создания экспертных систем (Colab, Python, библиотеки Keras, Numpy, фреймворк TensorFlow , система Clips)	Умеет интерпретировать результаты методов анализа данных и использовать их в научных исследованиях.
	ПКС-5.3. Владеет инструментами для сбора, обработки, интерпретации данных современных научных исследований	Имеет опыт обработки и интерпретации результатов анализа больших наборов данных с помощью программного обеспечения.

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 6 зачетных единиц, 216 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

Согласована на портале № 58037

Контактная аудиторная работа обучающихся с преподавателем, всего	72	72
Лекционные занятия	36	36
Лабораторные занятия	36	36
Самостоятельная работа обучающихся, в т.ч. контактная внеаудиторная работа обучающихся с преподавателем, всего	144	144
Подготовка к зачету	32	32
Подготовка к защите отчета по лабораторной работе	18	18
Подготовка к лабораторной работе, написание отчета	62	62
Написание отчета по лабораторной работе	26	26
Подготовка к тестированию	6	6
Общая трудоемкость (в часах)	216	216
Общая трудоемкость (в з.е.)	6	6

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					
1 Введение в аналитику больших данных	16	4	32	52	ПКС-3, ПКС-5, ОПК-2
2 Методы статистического анализа данных	10	8	34	52	ОПК-2, ПКС-3, ПКС-5
3 Методы машинного обучения	10	24	78	112	ОПК-2, ПКС-3, ПКС-5
Итого за семестр	36	36	144	216	
Итого	36	36	144	216	

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			

Согласована на портале № 58037

1 Введение в аналитику больших данных	Тема 1. Введение в анализ данных: понятие данных, типы шкал измерений, жизненный цикл данных, этапы решения задачи анализа данных и их взаимосвязи.	5	ПКС-3, ПКС-5
	Тема 2. Концепция больших данных: предпосылки формирования тренда, определение больших данных, источники больших данных, основные характеристики больших данных (4V), качество исходных данных, структурированность данных, драйверы рынка Big Data, функции и задачи больших данных, составляющие аналитики больших данных, процесс и принципы аналитики Big Data, отличия Big Data и Business intelligence (BI-аналитики), техники и методы анализа Big Data, специалисты Data Scientists.	6	ОПК-2, ПКС-5
	Тема 3. MapReduce и новый программный стек: потребности приложений для обработки Big Data, распределенные файловые системы, процесс вычислений в MapReduce, задачи-распределители, задачи-редукторы, группировка по ключу, комбинаторы, алгоритмы, в которых используется MapReduce, обобщения MapReduce для системы потоков работ, рекурсивные обобщения MapReduce, модель коммуникационной стоимости.	5	ОПК-2, ПКС-5
	Итого	16	
2 Методы статистического анализа данных	Тема 4. Основные задачи и методы статистического анализа, корреляционный анализ, дисперсионный анализ, регрессионный анализ, факторный анализ, кластерный анализ, дискриминантный анализ, многомерное шкалирование.	10	ОПК-2, ПКС-3, ПКС-5
	Итого	10	
3 Методы машинного обучения	Тема 5. Методы машинного обучения без учителя: кластеризация, сетевой анализ	5	ОПК-2, ПКС-3, ПКС-5
	Тема 5. Методы машинного обучения без учителя: кластеризация, сетевой анализ Тема 6. Методы машинного обучения с учителем: наивный Байесовский классификатор, комплексные модели, деревья решений, бустинг и бэггинг.	5	ОПК-2, ПКС-3, ПКС-5
	Итого	10	
Итого за семестр		36	
Итого		36	

5.3. Практические занятия (семинары)

Не предусмотрено учебным планом

Согласована на портале № 58037

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 Введение в аналитику больших данных	Шкалы измерений (расчет индекса экономики знаний)	4	ОПК-2, ПКС-3, ПКС-5
	Итого	4	
2 Методы статистического анализа данных	Регрессионные модели для предсказания поведения покупателей	8	ОПК-2, ПКС-3, ПКС-5
	Итого	8	
3 Методы машинного обучения	Кластерный анализ для сегментирования клиентской базы	6	ОПК-2, ПКС-3, ПКС-5
	Наивный байесовский классификатор для классификации твитов	6	ОПК-2, ПКС-3, ПКС-5
	Кластеризация на основе сетевых графов	6	ОПК-2, ПКС-3, ПКС-5
	Ансамбли моделей (бэггинг и бустинг)	6	ОПК-2, ПКС-3, ПКС-5
	Итого	24	
Итого за семестр		36	
Итого		36	

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 № 58037

1 Введение в аналитику больших данных	Подготовка к зачету	8	ОПК-2, ПКС-3, ПКС-5	Зачёт
	Подготовка к защите отчета по лабораторной работе	4	ОПК-2, ПКС-3, ПКС-5	Защита отчета по лабораторной работе
	Подготовка к лабораторной работе, написание отчета	12	ОПК-2, ПКС-3, ПКС-5	Лабораторная работа
	Написание отчета по лабораторной работе	6	ОПК-2, ПКС-3, ПКС-5	Отчет по лабораторной работе
	Подготовка к тестированию	2	ОПК-2, ПКС-3, ПКС-5	Тестирование
	Итого	32		
2 Методы статистического анализа данных	Подготовка к зачету	12	ОПК-2, ПКС-3, ПКС-5	Зачёт
	Подготовка к защите отчета по лабораторной работе	4	ОПК-2, ПКС-3, ПКС-5	Защита отчета по лабораторной работе
	Подготовка к лабораторной работе, написание отчета	12	ОПК-2, ПКС-3, ПКС-5	Лабораторная работа
	Написание отчета по лабораторной работе	4	ОПК-2, ПКС-3, ПКС-5	Отчет по лабораторной работе
	Подготовка к тестированию	2	ОПК-2, ПКС-3, ПКС-5	Тестирование
	Итого	34		
3 Методы машинного обучения	Подготовка к зачету	12	ОПК-2, ПКС-3, ПКС-5	Зачёт
	Подготовка к защите отчета по лабораторной работе	10	ОПК-2, ПКС-3, ПКС-5	Защита отчета по лабораторной работе
	Подготовка к лабораторной работе, написание отчета	38	ОПК-2, ПКС-3, ПКС-5	Лабораторная работа
	Написание отчета по лабораторной работе	16	ОПК-2, ПКС-3, ПКС-5	Отчет по лабораторной работе
	Подготовка к тестированию	2	ОПК-2, ПКС-3, ПКС-5	Тестирование
	Итого	78		
Итого за семестр		144		
Итого		144		

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

Согласована на портале № 58037

занятий

Формируемые компетенции	Виды учебной деятельности			Формы контроля
	Лек. зан.	Лаб. раб.	Сам. раб.	
ОПК-2	+	+	+	Зачёт, Защита отчета по лабораторной работе, Лабораторная работа, Тестирование, Отчет по лабораторной работе
ПКС-3	+	+	+	Зачёт, Защита отчета по лабораторной работе, Лабораторная работа, Тестирование, Отчет по лабораторной работе
ПКС-5	+	+	+	Зачёт, Защита отчета по лабораторной работе, Лабораторная работа, Тестирование, Отчет по лабораторной работе

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	10	10
Защита отчета по лабораторной работе	8	9	9	26
Лабораторная работа	8	9	9	26
Тестирование	4	4	4	12
Отчет по лабораторной работе	8	9	9	26
Итого максимум за период	28	31	41	100
Нарастающим итогом	28	59	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

Согласована на портале № 58037

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. Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д.У. Джеффри ; перевод с английского А.А. Слинкин. — Москва : ДМК Пресс, 2016. — 498 с. [Электронный ресурс]: — Режим доступа: https://e.lanbook.com/book/93571.

2. Прикладная математическая статистика: Учебное пособие / А. А. Мицель - 2019. 113 с. [Электронный ресурс]: — Режим доступа: https://edu.tusur.ru/publications/9151.

7.2. Дополнительная литература

1. Прикладная математическая статистика: Практические работы / А. А. Мицель - 2019. 81 с. [Электронный ресурс]: — Режим доступа: https://edu.tusur.ru/publications/9153.

2. Котиков, П. Е. Анализ данных : учебно-методическое пособие / П. Е. Котиков. — Санкт-Петербург : СПбГПМУ, 2019. — 48 с. [Электронный ресурс]: — Режим доступа: https://e.lanbook.com/book/174498.

7.3. Учебно-методические пособия

7.3.1. Обязательные учебно-методические пособия

1. Захарова А.А. Анализ данных в Excel и Calc: методические указания по выполнению лабораторных работ по дисциплине «Анализ больших данных» для направления обучения 09.04.01 «Информатика и вычислительная техника». – Томск: ТУСУР, 2022. – 76 с. [Электронный ресурс]: — Режим доступа: https://asu.tusur.ru/learning/090401e/d18/090401e-d18-labs.pdf.

2. Захарова А.А. Анализ больших данных: методические указания по самостоятельной работе студентов для направления 09.04.01 «Информатика и вычислительная техника» / А.А. Захарова. – Томск: ТУСУР, 2022. – 10 с. [Электронный ресурс]: — Режим доступа: https://asu.tusur.ru/learning/090401e/d18/090401e-d18-work.pdf.

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

Согласована на портале № 58037

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

Учебная вычислительная лаборатория / Лаборатория ГПО "Мониторинг": учебная аудитория для проведения занятий практического типа, учебная аудитория для проведения занятий лабораторного типа, помещение для курсового проектирования (выполнения курсовых работ), помещение для проведения групповых и индивидуальных консультаций, помещение для проведения текущего контроля и промежуточной аттестации, помещение для самостоятельной работы; 634034, Томская область, г. Томск, Вершинина улица, д. 74, 438 ауд.

Описание имеющегося оборудования:

- Рабочие станции: системный блок MB Asus P5B / CPU Intel Core 2 Duo 6400 2.13 GHz / 5Гб RAM DDR2 / 250Gb HDD / LAN (10 шт.);
- Монитор 19 Samsung 931BF (10 шт.);
- Проектор ACER X125H DLP;
- Экран проектора;
- Видеокамера (2 шт.);
- Точка доступа WiFi;
- Комплект специализированной учебной мебели;
- Рабочее место преподавателя.

Программное обеспечение:

- LibreOffice;
- Microsoft Office 2003;
- Microsoft Windows 7 Pro;

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, 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

Согласована на портале № 58037

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 Введение в аналитику больших данных	ПКС-3, ПКС-5, ОПК-2	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
		Отчет по лабораторной работе	Темы лабораторных работ
2 Методы статистического анализа данных	ОПК-2, ПКС-3, ПКС-5	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
		Отчет по лабораторной работе	Темы лабораторных работ

Согласована на портале № 58037

3 Методы машинного обучения	ОПК-2, ПКС-3, ПКС-5	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
		Отчет по лабораторной работе	Темы лабораторных работ

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
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

Согласована на портале № 58037

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 виды шкал относятся к качественным?
   А) наименований;
   Б) порядковая;
   В) интервалов;
   Г) отношений;
   Д) разностей;
   Е) абсолютная.
2. Основные характеристики больших данных (4V):
   А) объем, скорость создания и обработки, разнообразие источников и форм хранения, ценность;
   Б) объем, скорость создания и обработки, разнообразие источников и форм хранения, стоимость;
   В) структурированность, скорость создания и обработки, объем, ценность;
   Г) неструктурированность, скорость создания и обработки, объем, ценность
3. К какому типу информации относятся XML-документы и XSD-схемы?
   А) структурированная;
   Б) полуструктурированная;
   В) неструктурированная;
   Г) условно структурированная.
4. В чем состоит назначение задач-распределителей в технологии MapReduce?
   А) обрабатывают данные по одному ключу за раз и каким-то образом комбинируют значения, ассоциированные с этим ключом;
   Б) собирают значения, ассоциированные с каждым ключом, в список;
   В) преобразуют порцию в последовательность пар ключ-значение;
   Г) осуществляют сортировку по ключу.
5. Статистический метод анализа данных, предназначенный для исследования взаимозависимости выборок – это …
   А) кластерный анализ;
   Б) регрессионный анализ;
   В) корреляционный анализ;
   Г) факторный анализ.

Согласована на портале № 58037

6. Эти методы кластеризации последовательно объединяют, начиная с отдельных объектов, каждый раз во всё более крупные группы (кластеры) до тех пор, пока все объекты не попадают в один кластер.
7. Эти методы кластеризации последовательно объединяют, начиная с отдельных объектов, каждый раз во всё более крупные группы (кластеры) до тех пор, пока все объекты не попадают в один кластер.
   А) Иерархические дивизимные методы кластеризаци;
   Б) Иерархические агломеративные методы кластеризации;
   В) Неиерархические дивизимные методы кластеризации;
   Г) Неиерархические агломеративные методы кластеризации.
8. Если a – среднее расстояние от данного объекта до объектов из того же кластера, b – среднее расстояние от данного объекта до объектов из ближайшего кластера (отличного от того, в котором лежит сам объект), то силуэтом данного объекта называется величина:
   А) s=min(a,b).
   Б) s=(b−a)/min(a,b).
   В) s= max(a,b).
   Г) s=(b−a)/max(a,b).
9. С какой целью в наивном Байесовском классификаторе применяется логарифмирование?
   А) для решения проблемы арифметического переполнения;
   Б) для устранения проблемы появления неизвестных слов;
   В) для формирования вероятностного пространства;
   Г) для реализации bag of words model
10. Технология классификации, где все элементарные классификаторы обучаются и работают параллельно (независимо друг от друга).
    А) Бустинг;
    Б) Бэггинг;
    В) bag of words model;
    Г) наивный Байесовский классификатор.

9.1.2. Перечень вопросов для зачета

1. Понятие данных, информации, знаний
2. Типы шкал измерений
3. Жизненный цикл данных
4. Этапы решения задачи анализа данных и их взаимосвязи.
5. Предпосылки формирования тренда больших данных
6. Определение больших данных.
7. Источники больших данных
8. Основные характеристики больших данных (4V)
9. Проблема качества и структурированности исходных данных
10. Функции и задачи больших данных, составляющие аналитики больших данных
11. Процесс и принципы аналитики Big Data,
12. Отличия Big Data и Business intelligence (BI-аналитики),
13. Техники и методы анализа Big Data
14. Потребности приложений для обработки Big Data
15. Распределенные файловые системы
16. Процесс вычислений в MapReduce.
17. Модель коммуникационной стоимости.
18. Перечислите основные методы статистического анализа данных
19. Методы машинного обучения без учителя: кластеризация (виды и основные этапы)
20. Методы машинного обучения без учителя: сетевой анализ (сущность, основные понятия и этапы)
21. Методы машинного обучения с учителем: наивный Байесовский классификатор (сущность и этапы)
22. Методы машинного обучения с учителем: комплексные модели (ансамбли моделей), деревья решений,
23. Бустинг: понятие и суть метода
24. Бэггинг: понятие и суть метода

Согласована на портале № 58037

9.1.3. Примерный перечень вопросов для защиты лабораторных работ

1. Классификация методов кластеризации
2. Перечислите известные вам меры расстояния
3. Метрики качества кластеризации
4. Сущность метода k-средних
5. Сущность метода k-медиан
6. Понятие сетевого графа, вершины, ребра.
7. Матрица смежности
8. Степень вершины графа
9. Задача сетевого анализа
10. Понятие сообщества в сетевом анализе
11. Понятие случайного графа
12. Методы разделения сети на кластеры
13. Агломеративная и дивизионная кластеризация
14. Понятие модульности
15. Алгоритм «edge.betweenness.community»
16. Смысл теоремы Байеса.
17. Многомерная и мультиноминальная модель наивного Байесовского классификатора (NBC)
18. Для чего используется оценка апостериорного максимума в NBC?
19. Предположение условной независимости в NBC
20. Проблема арифметического переполнения в NBC
21. Оценка параметров Байесовской модели
22. Проблема неизвестных слов
23. Методика реализации наивного Байесовского классификатора
24. Формирование вероятностного пространства в NBC
25. Комплексные модели (ансамбли моделей)
26. Деревья решений,
27. Бустинг: понятие и суть метода
28. Бэггинг: понятие и суть метода
29. Алгоритм «Случайный лес»: сущность, достоинства и недостатки

9.1.4. Темы лабораторных работ

1. Шкалы измерений (расчет индекса экономики знаний)
2. Регрессионные модели для предсказания поведения покупателей
3. Кластерный анализ для сегментирования клиентской базы
4. Наивный байесовский классификатор для классификации твитов
5. Кластеризация на основе сетевых графов
6. Ансамбли моделей (бэггинг и бустинг)

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

Согласована на портале № 58037

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

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:

Согласована на портале № 58037

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 № 58037

ЛИСТ СОГЛАСОВАНИЯ

Рассмотрена и одобрена на заседании кафедры АСУ
протокол № 10 от «15» 10 2020 г.

СОГЛАСОВАНО:

Должность	Инициалы, фамилия	Подпись
Заведующий выпускающей каф. АСУ	В.В. Романенко	Согласовано, c3e2018f-3231-48c3-b093-89b6f5342191
Заведующий обеспечивающей каф. АСУ	В.В. Романенко	Согласовано, c3e2018f-3231-48c3-b093-89b6f5342191
Начальник учебного управления	Е.В. Саврук	Согласовано, fa63922b-1fce-4a6a-845d-9ce7670b004c

ЭКСПЕРТЫ:

Доцент, каф. АСУ	А.И. Исакова	Согласовано, 79bf1038-9d22-4279-a1e8-7806307b7f82
Заведующий кафедрой, каф. АСУ	В.В. Романенко	Согласовано, c3e2018f-3231-48c3-b093-89b6f5342191

РАЗРАБОТАНО:

Профессор, каф. АСУ	А.А. Захарова	Разработано, 1c4b3f34-d4dc-493c-800e-894835c5364f

Согласована на портале № 58037