

МИНОБРНАУКИ РОССИИ
ФГБОУ ВО «УФИМСКИЙ ГОСУДАРСТВЕННЫЙ
АВИАЦИОННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
ФАКУЛЬТЕТ ИНФОРМАТИКИ И РОБОТОТЕХНИКИ
Кафедра технической кибернетики

УТВЕРЖДАЮ:
Зав. кафедрой технической кибер-
нетики



О.Я. Бежаева
(подпись, инициалы, фамилия)
«01» сентября 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Обработка и анализ больших данных
(наименование дисциплины)

47.04.01 Философия
шифр и наименование направления подготовки (специальности)

направленность (профиль) подготовки «Философия искусственного интеллекта»
наименование направленности (профиля, специализации)

квалификация: магистр

форма обучения: очно-заочная
(очная, очно-заочная, заочная)

Разработчик программы, к.т.н., доцент

Насыров Рашит Вильевич

СОГЛАСОВАНО: *руководитель образовательной программы*
д.филос., наук, профессор БашГУ Елхова О.И.



Уфа – 2022

1. Цели освоения дисциплины

Формирование у студентов необходимой теоретической базы и практических навыков, которые позволят всесторонне и системно понимать современные проблемы прикладной математики и информатики, проблемы обработки и анализа информации, а также разрабатывать и анализировать концептуальные и теоретические модели при решении научных и прикладных задач в области информационных технологий. Одна из главных проблем современной обработки и анализа данных - рост объемов данных, поэтому вопросам обработки большого объема данных посвящена данная дисциплина. Главная задача курса - сформировать целостное представление о современных проблемах анализа и обработки больших данных, помочь овладеть опытом разработки и анализа концептуальных и теоретических моделей прикладных задач анализа больших данных с применением моделей Data Mining.

2. Место дисциплины в структуре образовательной программы

Цикл ООП

3. Результаты освоения.

УК-1 – Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели.

ОПК-1 – Способен применять в сфере своей профессиональной деятельности при решении нестандартных задач категории и принципы, характеризующие современные проблемы философии, предлагать и аргументированно обосновывать способы их решения.

ОПК-3 – Способен использовать в сфере своей профессиональной деятельности категории и принципы методики преподавания философии и педагогики высшей школы.

Наименование категории (группы) общепрофессиональных компетенций	Код и наименование профессиональной компетенции	Код и наименование индикатора достижения профессиональной компетенции
<i>Командная работа и лидерство</i>	УК-3 – способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения по-	ИУК 3.1. <i>Знать:</i> – общие формы организации деятельности коллектива; – психологию межличностных отношений в группах разного возраста; – основы стратегического планирования работы коллектива для достижения поставленной цели. ИУК 3.2. <i>Уметь:</i> – создавать в коллективе психологически безопасную доброжелательную среду; – учитывать в своей социальной и профессио-

	<p>ставленной цели</p>	<p>нальной деятельности интересы коллег;</p> <ul style="list-style-type: none"> – предвидеть результаты (последствия) как личных, так и коллективных действий; – планировать командную работу, распределять поручения и делегировать полномочия членам команды. <p>ИУК 3.3. Владеть:</p> <ul style="list-style-type: none"> – навыками постановки цели в условиях командой работы; – способами управления командной работой в решении поставленных задач; – навыками преодоления возникающих в коллективе разногласий, споров и конфликтов на основе учета интересов всех сторон.
<p><i>Профессиональные исследования</i></p>	<p>ОПК-1 – Способен применять в сфере своей профессиональной деятельности при решении нестандартных задач категории и принципы, характеризующие современные проблемы философии, предлагать и аргументированно обосновывать способы их решения.</p>	<p>ИОПК 1.1. Знать:</p> <ul style="list-style-type: none"> – категории и принципы, характеризующие современные проблемы философии; – основные методы аргументации и обоснования принятия решения. <p>ИОПК 1.2. Уметь:</p> <ul style="list-style-type: none"> – применять в сфере своей профессиональной деятельности при решении нестандартных задач категории и принципы, характеризующие современные проблемы философии, основные методы аргументации и обоснования принятия решения. <p>ИОПК 1.3. Владеть:</p> <ul style="list-style-type: none"> – категориями и принципами, характеризующими современные проблемы философии, основными методами аргументации и обоснования
<p><i>Профессиональная коммуникация</i></p>	<p>ОПК-3 – Способен использовать в сфере своей профессиональной деятельности</p>	<p>ИОПК 3.1. Знать:</p> <ul style="list-style-type: none"> – категории и принципы осуществления профессиональной коммуникации по профилю деятельности в мультикультурной среде. <p>ИОПК 3.2. Уметь:</p> <ul style="list-style-type: none"> – применять в сфере своей профессиональ-

	<p>категории и принципы методики преподавания философии и педагогики высшей школы.</p>	<p>ной деятельности категории и принципы методики преподавания философии и педагогики высшей школы</p> <p>ИОПК 3.3. <i>Владеть:</i></p> <ul style="list-style-type: none"> – способами использования в сфере своей профессиональной деятельности категорий и принципов методики преподавания философии и педагогики высшей школы
--	--	---

В результате освоения дисциплины обучающийся должен

4.1. Знать:

- 4.1.1** • методы решения задач обработки и анализа больших данных, возможности высокопроизводительных вычислительных систем, технологии распределенных вычислений, методы и модели Data Mining.

4.2. Уметь:

- разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных;
- использовать и применять углубленные знания в области обработки и анализа

4.2.1 больших данных;

- оценивать время и необходимые аппаратные ресурсы для решения задач анализа и обработки данных;
- создавать алгоритмы анализа и обработки большого объема данных с применением моделей Data Mining.

4.3. Иметь навыки и (или) опыт деятельности (владеть):

- 4.3.1** • навыками применения программных систем, предназначенных для анализа больших данных.

5. Тематический план учебной дисциплины

Курс рассчитан на 48 часов аудиторной нагрузки, из них 18 часов лекций, 12 часов лабораторных работ, 6 часов практических занятий, общим объемом 4 зачетные единицы (144 часа). Выходной контроль - экзамен

5.1. Структура

№	Название раздела	Всего часов	Аудиторные часы			Самостоятельная работа
			Лекции	Лабораторные работы	Практические занятия	
1. Технологии анализа данных (1 семестр)						
1	Большие данные (Big Data):	10	2			8
2	Процесс анализа.	11	2	4	2	6
3	Технологии KDD и Data Mining.	11	2	4		6
4	Программное обеспечение в области анализа данных.	11	2	4		6
2. Интеллектуальный анализ данных (2 семестр)						
5.	Ассоциативные правила.	10	2		2	8
6.	Определение кластеризации.		2		2	8
7.	Применение классификации и регрессии.		2	4		6
8.	Основные понятия теории нейронных сетей.		2	4		6
9.	Определение дерева решений.		2	4		6
	Итого:	108	18	24	6	60

Лекции (1семестр)

№ п/п	№ раздела	Тема	Контролируемые действия по проверке ЗУВ1	часы	в том числе в форме практической подготовки, час	Оценочное средство ²	Критерии оценки				
							отлично	хорошо	Удовлетворительно	Неудовлетворительно	
Без использования электронных образовательных ресурсов											
1	2	Большие данные (Big Data): современные подходы к обработке и хранению. Проблема множественного сравнения данных.		2							
2	2	Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных.		2							
3	2	Технологии KDD и Data Mining. Подготовка данных к анализу. Методика извлечения знаний. Data Mining. Мультидисциплинарный характер Data Mining. Причины распространения KDD и Data Mining. Актуальность технологий Data Mining		2							

¹ Совокупный набор контролируемых действий по проверке ЗУВ, указанных в разделе 3 без повторений, должен совпадать с набором, указанным в разделе 1.

² Примерный перечень возможных оценочных средств: деловая/ролевая игра, кейс-задача, контрольная работа, круглый стол, дискуссия, портфолио, проект, рабочая тетрадь, задачи, задания, реферат, доклад, сообщение, собеседование, тест, тренажёр, эссе, творческое задание.

		как средств обработки больших объемов информации.						
4	2	Программное обеспечение в области анализа данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования.		2				

Лекции (2семестр)

№ п/п	№ раздела	Тема	Контролируемые действия по проверке ЗУВЗ	часы	в том числе в форме практической подготовки, час	Оценочное средство ⁴	Критерии оценки			
							отлично	хорошо	Удовлетворительно	Неудовлетворительно
Без использования электронных образовательных ресурсов										
1	2	Ассоциативные правила. Аффинитивный анализ, предметный набор. Поддержка и достоверность ассоциативного правила. Значимость ассоциативных правил, лифт и леведж. Поиск ассоциативных правил. Частые предметные наборы и их обнаружение. Алгоритм генерации ассоциативных правил. Иерархические ассоциативные правила. Методы поиска иерархи-		2						

³ Совокупный набор контролируемых действий по проверке ЗУВ, указанных в разделе 3 без повторений, должен совпадать с набором, указанным в разделе 1.

⁴ Примерный перечень возможных оценочных средств: деловая/ролевая игра, кейс-задача, контрольная работа, круглый стол, дискуссия, портфолио, проект, рабочая тетрадь, задачи, задания, реферат, доклад, сообщение, собеседование, тест, тренажёр, эссе, творческое задание.

		ческих ассоциативных правил.						
2	2	Определение кластеризации. Постановка задачи кластеризации. Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Меры расстояний. Пример работы алгоритма k-means. Проблемы алгоритмов кластеризации.		2				
3	2	Применение классификации и регрессии. Обзор методов классификации и регрессии. Статистические методы. Методы, основанные на обучении, разнообразие подходов.		2				
4	2	Основные понятия теории нейронных сетей. Основные парадигмы нейронных сетей. Многослойный персептрон: класс решаемых задач, архитектура.		2				
5	2	Определение дерева решений. Причины популярности и условия применимости. Структура дерева решений. Выбор атрибута разбиения в узле. Алгоритм ID3, критерий выбора атрибута разбиения ID3, пример работы алгоритма. Проблема пе-		2				

		реобучения, Неизвестные значения атрибутов, алгоритм С4.5.							
--	--	--	--	--	--	--	--	--	--

Лабораторные работы (1 семестр)

№ п/п	№ раз-дела	Наименование	Контролируемые действия по проверке ЗУВ	часы	в том числе в форме практической подготовки, час	Оценочное средство	Критерии оценки			
							отлично	хорошо	удовлетворительно	неудовлетворительно
1	2	Процесс анализа.		4						
2	3	Технологии KDD и Data Mining.		4						
3	4	Программное обеспечение в области анализа данных.		4						

Лабораторные работы (2 семестр)

№ п/п	№ раз-дела	Наименование	Контролируемые действия по проверке ЗУВ	часы	в том числе в форме практической подготовки, час	Оценочное средство	Критерии оценки			
							отлично	хорошо	удовлетворительно	неудовлетворительно
							зачтено			не зачтено
		Прогнозирование с помощью линейной регрессии.		4						
		Классификация с помощью нейросети.		4						
		Классификация с помощью деревьев решений.		4						

Практические занятия (1 семестр)

№ п/п	№ раз-дела	Тема	Контролируемые действия по проверке ЗУВ	часы	в том числе в форме практической подготовки, час	Оценочное средство	Критерии оценки			
							отлично	хорошо	удовлетворительно	неудовлетворительно
							зачтено			не зачтено
Без использования электронных образовательных ресурсов										
1	2	Понятие сценария и узла обработки. Консолидация данных. Трансформация данных. Визуализация данных.		2						

Практические занятия (21 семестр)

№ п/п	№ раз-дела	Тема	Контролируемые действия по проверке ЗУВ	часы	в том числе в форме практической подготовки, час	Оценочное средство	Критерии оценки			
							отлично	хорошо	удовлетворительно	неудовлетворительно
							зачтено			не зачтено
Без использования электронных образовательных ресурсов										
1		Ассоциативные правила. Поиск ассоциативных правил.		2						
2		Кластеризация. Алгоритм кластеризации k-means.		2						

Самостоятельное изучение разделов

№ п/п	№ раз-дела	Темы и вопросы	Контролируемые действия по проверке ЗУВ	часы	в том числе в форме практической	Оценочное средство	Критерии оценки			
							отлично	хорошо	удовлетворительно	неудовлетворительно
							зачтено			не зачтено

					подготовки, час			
		Без использования электронных образовательных ресурсов						
		Проработка необходимых вопросов для подготовки к лекциям и практическим занятиям.						

7. Примеры заданий на КР

Примеры заданий контрольной работы

1. Составить регулярное выражение, удовлетворяющее заданным требованиям.
2. Построить наиболее вероятную цепочку тегов (скрытых состояний) в заданной скрытой марковской модели по указанному предложению.
3. Вывести формулу для коэффициентов заданного алгоритма сглаживания n-граммной языковой модели.
4. Построить символьную триграммную языковую модель по заданному корпусу и с ее помощью построить распознаватель языка документа.
5. Вычислить перплексию n-граммной языковой модели с заданным сглаживанием.
6. На основе заданной обучающей выборки построить марковскую модель максимальной энтропии для выделения заданных именованных сущностей (имен собственных, географических названий и т. д.) из текста.

8 Образовательные технологии

Основными образовательными технологиями являются: работа в группах на лабораторных и практических занятиях, проектный метод.

9 Учебно-методическое и информационное обеспечение дисциплины

9.1. Рекомендуемая литература

9.1.1. Основная литература

	Авторы	Заглавие	Издательство, год	Эл. адрес
Л1.1	Мхитарян В.С. - Отв. ред.	АНАЛИЗ ДАННЫХ. Учебник для академического бакалавриата: Гриф УМО ВО	М.:Издательство Юрайт, 2018	https://biblio-online.ru/book/CC38E97A-CCE5-4470-90F1-3B6D35ACC0B4
Л1.2	Боровков А.А.	Математическая статистика: учебник	СПб.: Лань, 2010	https://e.lanbook.com/book/3810
Л1.3	Жуковский, О.И.	Информационные технологии и анализ данных : учебное пособие	Томск : Эль Контент, 2014	http://biblioclub.ru/index.php?page=book&id=480500

9.1.2. Дополнительная литература

	Авторы	Заглавие	Издательство, год	Эл. адрес
Л2.1	Симчера В. М.	Методы многомерного анализа статистических данных:	Финансы и статистика, 2008	http://biblioclub.ru/index.php?page=book&id=59559&sr=1

		учебное пособие		
Л2.2	Вирт Н.	Алгоритмы и структуры данных. Новая версия для Оберона: Учебные пособия	Издательство "ДМК Пресс", 2010	https://e.lanbook.com/book/1261

9.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

	Название	Эл. адрес
Э1	Курс в Moodle "Обработка и анализ больших данных"	https://sdo.ugatu.ru/

9.3 Ресурсы информационно-телекоммуникационной сети «Интернет»

Информационная справочная система:

СПС КонсультантПлюс (<http://www.consultant.ru/>).

Профессиональные базы данных:

1. Электронная база данных «Scopus» (<http://www.scopus.com>);
2. Электронная библиотечная система Алтайского государственного университета (<http://elibrary.asu.ru/>);
3. Научная электронная библиотека elibrary (<http://elibrary.ru>).
4. Электронная база данных «ZBMATH – The database Zentralblatt MATH» <https://zbmath.org/>
5. 7. Электронная база данных ZBMATH: <https://zbmath.org/>

10 Рекомендации для самостоятельной работы студентов

Самостоятельная работа может рассматриваться как организационная форма обучения – система педагогических условий, обеспечивающих управление учебной деятельностью по освоению знаний и умений в области учебной деятельности без посторонней помощи. Студенту нужно четко понимать, что самостоятельная работа – не просто обязательное, а необходимое условие для получения знаний по дисциплине и развитию компетенций, необходимых в будущей профессиональной деятельности.

Самостоятельная работа проводится с целью:

- систематизации и закрепления полученных на лекциях теоретических знаний;
- углубления и расширения теоретических знаний;
- формирования умений использовать нормативную, правовую, справочную документацию и специальную литературу;
- развития познавательных способностей и активности студентов: творческой инициативы, самостоятельности, ответственности и организованности;
- формирования самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;
- формирования практических (общих и профессиональных) умений и навыков;
- развития исследовательских умений;

– получения навыков эффективной самостоятельной профессиональной (практической и научно-теоретической) деятельности.

11 Материально-техническое обеспечение дисциплины и информационные технологии, используемые при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения информационных справочных систем (при необходимости).

Практические занятия проводятся в компьютерных классах. На лекциях и практических занятиях используется проектор. Для успешного освоения дисциплины, студент использует следующие программные средства: Python (пакеты `scipy` и `numpy`, сборка `Anaconda`, `Pandas`, `Scikit-learn` и др.), инструменты R и RStudio.

12 Особенности организации обучения для лиц с ограниченными возможностями здоровья

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) могут предлагаться следующие варианты восприятия учебной информации с учетом их индивидуальных психофизических особенностей, в том числе с применением электронного обучения и дистанционных технологий:

1) для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат);

индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания и консультации.

2) для лиц с нарушениями слуха: в печатной форме; в форме электронного документа;

видеоматериалы с субтитрами; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации.

3) для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.