


МИНОБРНАУКИ РОССИИ
АСТРАХАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМ. В.Н. ТАТИЩЕВА

СОГЛАСОВАНО
Руководитель ОПОП


И.И. Гордеев
29 июня 2022 г.

УТВЕРЖДАЮ
Заведующий кафедрой ПМИ


М.В. Коломина
29 июня 2022 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
АНАЛИЗ И ОБУЧЕНИЕ НА БОЛЬШИХ ДАННЫХ**

Составитель(-и)	Русак А. В., к.т.н., ИТМО Коломина М.В., к.ф.м.н., доцент каф. ПМИ, АГУ Кузнецова В.Ю., ассистент каф.ИБ, АГУ
Направление подготовки / специальность	09.04.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ
Направленность (профиль) ОПОП	ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА
Квалификация (степень)	магистр
Форма обучения	очная
Год приема	2022
Курс	2

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1. Целью освоения дисциплины «Анализ и обучение на больших данных» является формирование у студентов компетенций в области анализа и обучения на больших данных.

1.2. Задачи освоения дисциплины (модуля):

- изучение видов и методов представления данных и их структур;
- приобретение практических навыков применения методов машинного обучения на больших данных;
- умение проектировать и администрировать базы данных.

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОПОП

2.1. Учебная дисциплина Б1.В.Д.01.02 «Анализ и обучение на больших данных» относится к части, формируемой участниками образовательных отношений учебного плана направления подготовки 09.04.02 Информационные системы и технологии 2022 года набора.

2.2. Для изучения данной учебной дисциплины (модуля) необходимы следующие знания, умения и навыки, формируемые предшествующими дисциплинами:

- Специальные главы математики;
- Прикладной искусственный интеллект;
- Обработка и анализ данных.

2.3. Перечень последующих учебных дисциплин, для которых необходимы знания, умения и навыки, формируемые данной учебной дисциплиной:

- Производственная практика;
- Выпускная квалификационная работа (магистерская диссертация).

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

Процесс изучения дисциплины направлен на формирование элементов следующих компетенций в соответствии с ФГОС ВО и ОПОП ВО по данному направлению подготовки:

а) профессиональных (ПК): ПК-14. Способен осуществлять руководство по созданию и развитию систем и комплексов обработки данных, в том числе больших данных, для корпоративных и государственных заказчиков

**Таблица 1.
Декомпозиция результатов обучения**

Код компетенции, индикатора	Планируемые результаты освоения дисциплины (модуля)		
	Знать (1)	Уметь (2)	Владеть (3)
ПК-14 ПК-14.1. Осуществляет личное участие в проектах в роли архитектора центра обработки данных, технологического эксперта, специалиста по предпродажам.	ПК-14.1.1 существующие и перспективные структуры центров обработки данных; действующую нормативную базу в области проектирования и строительства центров обработки данных.	ПК-14.1.2 формулировать технические задания по формированию аппаратного обеспечения и программных комплексов центра обработки данных; вести переговоры с подрядчиками и поставщиками оборудования и программного обеспечения для центра обработки данных.	ПК-14.1.3 навыками составления технического задания и ведения переговоров.
ПК-14 ПК-14.2. Планирует и осуществляет технологическое развитие центров обработки данных, наращивание и поддержание технологических мощно-	ПК-14.2.1 принципы разработки проектной документации центра обработки данных; системную архитектуру серверного оборудования и систем хранения данных, циф-	ПК-14.2.2 ставить задачи и планировать работу сотрудников центра обработки данных.	ПК-14.2.3 навыками постановки задач и планирования работ по технологическому развитию центров обработки данных; организации эффек-

стей и компетенций подразделений.	ровых платформ анализа данных; методы работы с UNIX-подобными системами.		тивной команды.
ПК-14 ПК-14.3. Участвует в создании (модернизации) общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных; хранения наборов данных (в том числе звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения) на общедоступных платформах для обеспечения потребностей организаций-разработчиков в области искусственного интеллекта.	ПК-14.3.1 принципы и методы построения общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных; принципы и методы хранения наборов данных (в том числе звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения) на общедоступных платформах для обеспечения потребностей организаций-разработчиков в области искусственного интеллекта.	ПК-14.3.2 применять принципы и методы построения общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных; применять принципы и методы хранения наборов данных (в том числе, звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения) на общедоступных платформах для обеспечения потребностей организаций-разработчиков в области искусственного интеллекта.	ПК-14.3.3 навыками применения принципов хранения различных наборов данных на общедоступных платформах.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Объем дисциплины (модуля) в зачетных единицах **3 з. е.** Всего 108 часов: 54 часа выделено на контактную работу обучающихся с преподавателем (лекции – 18, лабораторные работы – 36), 54 часа – на самостоятельную работу обучающихся:

Таблица 2.

Структура и содержание дисциплины (модуля)

№ п/п	Наименование раздела (темы)	Семестр	Неделя семестра	Контактная работа (в часах)			Самостоят. работа		Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
				Л	ПЗ	ЛР	КР	СР	
1	Введение в проблематику построения систем анализа больших данных	3	1-4	2		0		8	Кейс 1
2	Методы и технологии анализа больших данных		5-10	8		20		28	Кейс 2, 3 Лабораторные работы 1-4
3	Обучение на больших данных		11-18	8		16		18	Лабораторные работы 5-6
ИТОГО				18		36		54	ЭКЗАМЕН

Условные обозначения:

Л – занятия лекционного типа; ПЗ – практические занятия, ЛР – лабораторные работы; КР – курсовая работа; СР – самостоятельная работа по отдельным темам.

Таблица 3.

Матрица соотношения тем/разделов учебной дисциплины/модуля и формируемых в них компетенций

Темы, Разделы дисциплины	Кол-во часов	Компетенции	
		ПК 14	общее количество компетенций
Введение в проблематику построения систем анализа больших данных	10	+	1

Методы и технологии анализа больших данных	56	+	1
Обучение на больших данных	42	+	1
Итого	108		

Краткое содержание дисциплины

Тема 1. Введение в проблематику построения систем анализа больших данных.

Понятие и классификация больших данных. Размерность и объем данных. «Проклятие размерности». Сценарии применения технологий больших данных в различных отраслях. Жизненный цикл аналитики данных, методологии CRISP-DM и SEMMA. Компетенции и состав команды для работы с данными. Правовые аспекты организации защиты персональных данных. Общий регламент защиты персональных данных (GDPR).

Тема 2. Методы и технологии анализа больших данных

Задачи и методы извлечения знаний из больших данных (Data Mining). Инструменты Data Mining. Методы сбора и предобработки данных из различных источников. Примеры применения Data Mining для различных отраслей. Введение в анализ социальных сетей. Базовые алгоритмы на графах и основные возможности графового анализа. Инструменты и методы визуализации графов. Средства визуализации для аналитики данных. Инструментарий для работы с большими данными. Основы работы и функционал компонентов экосистемы Hadoop, парадигма MapReduce. Apache Spark и его компоненты. Обзор облачных платформ для работы с большими данными. Масштабирование и многоуровневое хранение данных: Парадигма NoSQL.

Тема 3. Обучение на больших данных

Особенности представления и обработки символьной и численной информации в нейронных сетях. Методы и средства аннотирования данных для обучения нейронных сетей. Современные архитектуры нейронных сетей: сверточные нейронные сети и автокодировщики, разновидности рекуррентных нейронных сетей, состязательные сети, трансформеры, графовые нейронные сети. Современные алгоритмы и принципы обучения нейронных сетей, особенности обучения сетей различной структуры. Подходы к решению типовых задач, в том числе в области программной инженерии с использованием ИС. Современные инструментальные средства для разработки и обучения моделей искусственных нейронных сетей. Анализ точности построения и валидация (оценка практической применимости) нейросетевой модели. Системы нечеткого вывода: нечеткие высказывания, вывод в нечеткой логике, правила нечетких продукций. Нечеткие нейронные (гибридные) сети: определение, виды архитектур, обучение и оптимизация.

5. ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

5.1. Указания по организации и проведению лекционных, практических (семинарских) и лабораторных занятий с перечнем учебно-методического обеспечения.

Лекционные занятия

Основной формой реализации теоретического обучения является лекция, которая представляет собой систематическое, последовательное изложение преподавателем-лектором учебного материала теоретического характера. Цель лекции – организация целенаправленной познавательной деятельности студентов по овладению программным материалом учебной дисциплины.

Порядок подготовки лекционного занятия включает в себя выполнение следующих этапов:

- изучение требований программы дисциплины;
- определение целей и задач лекции;
- разработка плана проведения лекции;
- подбор литературы (ознакомление с методической литературой, публикациями периодической печати по теме лекционного занятия);
- отбор необходимого и достаточного по содержанию учебного материала;
- определение методов, приемов и средств поддержания интереса, внимания, стимулирования творческого мышления студентов;

- написание конспекта лекции.
Лекция должна включать следующие разделы:
- формулировку темы лекции;
- указание основных изучаемых разделов или вопросов и предполагаемых затрат времени на их изложение;
- изложение вводной части;
- изложение основной части лекции;
- краткие выводы по каждому из вопросов;
- заключение;
- рекомендации литературных источников по излагаемым вопросам.

Лабораторные занятия

Лабораторное занятие – целенаправленная форма организации педагогического процесса, направленная на углубление научно-теоретических знаний и овладение определенными методами работы, в процессе которых вырабатываются умения и навыки выполнения тех или иных учебных действий в данной сфере науки. Они развивают научное мышление и речь, позволяют проверить знания студентов и выступают как средства оперативной обратной связи.

Правильно организованные лабораторные занятия ориентированы на решение следующих задач:

- обобщение, систематизация, углубление, закрепление полученных на лекциях и в процессе самостоятельной работы теоретических знаний по дисциплине (предмету);
- формирование практических умений и навыков, необходимых в будущей профессиональной деятельности, реализация единства интеллектуальной и практической деятельности;
- выработка при решении поставленных задач таких профессионально значимых качеств, как самостоятельность, ответственность, точность, творческая инициатива.

Состав заданий для лабораторного занятия должен быть спланирован с расчетом, чтобы за отведенное время они могли быть качественно выполнены большинством учащихся.

Лабораторные занятия должны так быть организованы, чтобы студенты ощущали нарастающее сложность выполнения заданий, испытывали бы положительные эмоции от переживания собственного успеха в учении, поисками правильных и точных решений.

Самостоятельная работа

Самостоятельная работа – это вид учебной деятельности, которую студент совершает в установленное время и в установленном объеме индивидуально или в группе, без непосредственной помощи преподавателя (но при его контроле), руководствуясь сформированными ранее представлениями о порядке и правильности выполнения действий.

В учебном процессе образовательного учреждения выделяются два вида самостоятельной работы:

- аудиторная – выполняется на учебных занятиях, под непосредственным руководством преподавателя и по его заданию (выполнение самостоятельных работ; выполнение контрольных и практических работ; решение задач);
- внеаудиторная – выполняется по заданию преподавателя, но без его непосредственного участия (подготовка к аудиторным занятиям; изучение учебного материала, вынесенного на самостоятельную проработку; выполнение домашних заданий разнообразного характера; выполнение индивидуальных заданий, направленных на развитие у студентов самостоятельности и инициативы; подготовка к контрольной работе). Внеаудиторные самостоятельные работы представляют собой логическое продолжение аудиторных занятий, проводятся по заданию преподавателя, который инструктирует студентов и устанавливает сроки выполнения задания.

5.2. Указания для обучающихся по освоению дисциплины (модулю)

Лекция

- Лекция – основной вид обучения в вузе.

- В лекции излагаются основные положения теории, ее понятия и законы, приводятся факты, показывающие связь теории с практикой.
- Накануне лекции необходимо повторить содержание предыдущей лекции (а также теорию по изучаемой теме в школьных учебниках геометрии, если эта тема была представлена в них), а затем посмотреть тему очередной лекции по программе (по плану лекций).
- Полезно вести записи (конспекты) лекций: для непонятных вопросов оставлять место при работе над темой лекции с учебными пособиями.
- Записи лекций следует вести в отдельной тетради, оставляя место для дополнений во время самостоятельной работы.
- При конспектировании лекций выделяйте главы и разделы, параграфы, подчеркивайте основное.

Лабораторное занятие

- Лабораторное занятие – наиболее активный вид учебных занятий в вузе. Он предполагает самостоятельную работу над лекциями и учебными пособиями.
- К каждому лабораторному занятию нужно готовиться. Подготовку следует начинать с повторения теории (по записям лекций или по учебному пособию). После этого нужно решать задачи из предложенного домашнего задания.

Организация самостоятельной работы

Самостоятельность в учебной работе способствует развитию заинтересованности студента в изучаемом материале, вырабатывает у него умение и потребность самостоятельно получать знания, что весьма важно для специалиста с высшим образованием. Самостоятельная работа студентов представлена в следующих формах:

- работа с учебной литературой и конспектом лекций с целью подготовки к лабораторным занятиям, составление конспектов тем, выносимых на самостоятельную проработку;
- систематическое выполнение домашних работ.

Таблица 4.
Содержание самостоятельной работы обучающихся

Номер радела (темы)	Темы/вопросы, выносимые на самостоятельное изучение	Кол-во часов	Формы работы
1	Введение в проблематику построения систем	8	Изучение теоретического материала.
2	Методы и технологии анализа больших данных	28	Изучение теоретического материала. Подготовка к лабораторным работам.
3	Обучение на больших данных	18	Изучение теоретического материала. Подготовка к лабораторным работам.

5.3. Виды и формы письменных работ, предусмотренных при освоении дисциплины, выполняемые обучающимися самостоятельно.

Отчет по лабораторной работе – оформляется и отчитывается в электронном виде: формат листа А4, книжная ориентация страницы. Отчеты по всем лабораторным работам имеют единый титульный лист, на котором указывается наименование дисциплины, ФИО и группа исполнителя, ФИО преподавателя, принимающего отчеты. В отчете по каждой лабораторной работе должно быть представлено наименование работы, цель, ход выполнения работы (скриншоты, краткое текстовое описание), выводы по результатам работы.

6. ОБРАЗОВАТЕЛЬНЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

При реализации различных видов учебной работы по дисциплине могут использоваться электронное обучение и дистанционные образовательные технологии.

6.1. Образовательные технологии

Учебные занятия по дисциплине могут проводиться с применением информационно-телекоммуникационных сетей при опосредованном (на расстоянии) интерактивном взаимодействии обучающихся и преподавателя в режимах on-line или off-line в формах.

№	Формы	Описание
1	Лекция-дискуссия	Лекция-дискуссия специально не назначается, а возникает достаточно спонтанно на большинстве лекций. Студенты устно высказывают своё мнение по ходу лекции, дискутируют как с лектором, так и между собой. Также дискуссии иногда возникают при защите лабораторных работ при обсуждении выбранных особенностей реализации программного кода.
2	Исследовательские методы в обучении	Дает возможность учащимся самостоятельно пополнять свои знания, глубоко вникать в изучаемую проблему и предполагать пути ее решения, что важно при формировании мировоззрения.
3	Лабораторные работы	Формирование навыков использования современных компьютерных технологий по отладке программ.
4	Самостоятельная работа	Работа с ресурсами Internet, подготовка к лабораторным работам.

6.2. Информационные технологии

При реализации различных видов учебной и внеучебной работы используются следующие информационные технологии:

- система управления обучением LMS Moodle;
- использование возможностей Интернета в учебном процессе (рассылка заданий, предоставление выполненных работ, ответы на вопросы, ознакомление обучающихся с оценками и т.д.);
- использование электронных учебников и различных сайтов (например, электронные библиотеки, журналы и т.д.) как источник информации;
- использование возможностей электронной почты;
- использование средств представления учебной информации (электронных учебных пособий, применение новых технологий для проведения занятий с использованием презентаций и т.д.);
- использование интерактивных средств взаимодействия участников образовательного процесса (технологии дистанционного или открытого обучения в глобальной сети);
- использование интегрированных образовательных сред, где главной составляющей являются не только применяемые технологии, но и содержательная часть, т.е. информационные ресурсы (доступ к мировым информационным ресурсам, на базе которых строится учебный процесс).

6.3. Перечень программного обеспечения и информационных справочных систем

а) Перечень лицензионного учебного программного обеспечения:

Наименование программного обеспечения	Назначение
Adobe Reader	Программа для просмотра электронных документов
Платформа дистанционного обучения LMS Moodle	Виртуальная обучающая среда
Google Chrome	Браузер
Microsoft Office 2013, Microsoft Office Project 2013, Microsoft Office Visio 2013	Офисная программа
7-zip	Архиватор
Microsoft Windows 7 Professional	Операционная система
Kaspersky Endpoint Security	Средство антивирусной защиты
Notepad++	Текстовый редактор
PyCharm EDU	Среда разработки
R	Программная среда вычислений
Scilab	Пакет прикладных математических программ

Sofa Stats	Программное обеспечение для статистики, анализа и отчетности
VirtualBox	Программный продукт виртуализации операционных систем
VMware (Player)	Программный продукт виртуализации операционных систем
Microsoft Visual Studio	Среда разработки
Oracle SQL Developer	Среда разработки
IBM SPSS Statistics 21	Программа для статистической обработки данных

б) Современные профессиональные базы данных и информационные справочные системы:

1. Электронный каталог Научной библиотеки АГУ на базе MARK SQL НПО «Информ-систем»: <https://library.asu.edu.ru>.
2. Электронный каталог «Научные журналы АГУ»: <http://journal.asu.edu.ru/>.
3. Универсальная справочно-информационная полнотекстовая база данных периодических изданий ООО «ИВИС»: <http://dlib.eastview.com/>
4. Электронно-библиотечная система eLibrary. <http://elibrary.ru>
5. Справочная правовая система КонсультантПлюс: <http://www.consultant.ru>
6. Информационно-правовое обеспечение «Система ГАРАНТ»: <http://garant-astrakhan.ru>

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

7.1. Паспорт фонда оценочных средств

При проведении текущего контроля и промежуточной аттестации по дисциплине «Анализ и обучение на больших данных» проверяется сформированность у обучающихся компетенций, указанных в разделе 3 настоящей программы. Этапность формирования данных компетенций в процессе освоения образовательной программы определяется последовательным освоением дисциплин (модулей) и прохождением практик, а в процессе освоения дисциплины – последовательным достижением результатов освоения содержательно связанных между собой разделов, тем.

Таблица 5

Соответствие разделов, тем дисциплины (модуля), результатов обучения по дисциплине (модулю) и оценочных средств

№ п/п	Контролируемые разделы, темы дисциплины (модуля)	Код контролируемой компетенции (компетенций)	Наименование оценочного средства
1.	Введение в проблематику построения систем анализа больших данных	ПК-14	Кейс 1
2.	Методы и технологии анализа больших данных	ПК-14	Кейс 2, 3. Лабораторные работы 1-4
3.	Обучение на больших данных	ПК-14	Лабораторные работы 5-6

7.2. Описание показателей и критериев оценивания компетенций, описание шкал оценивания

Для оценки результатов обучения применяются следующие критерии:

Таблица 6

Показатели оценивания результатов обучения в виде знаний

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует глубокое знание теоретического материала, умение обоснованно излагать свои мысли по обсуждаемым вопросам, способность полно, правильно и аргументированно отвечать на вопросы, приводить примеры
4 «хорошо»	демонстрирует знание теоретического материала, его последовательное изложение, способность приводить примеры, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует неполное, фрагментарное знание теоретического материала, требующее наводящих вопросов преподавателя, допускает существенные ошибки в его изложении,

	затрудняется в приведении примеров и формулировке выводов
2 «неудовлетворительно»	демонстрирует существенные пробелы в знании теоретического материала, не способен его изложить и ответить на наводящие вопросы преподавателя, не может привести примеры

Таблица 7

Показатели оценивания результатов обучения в виде умений и владений

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы
4 «хорошо»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует отдельные, несистематизированные навыки, не способен применить знание теоретического материала при выполнении заданий, испытывает затруднения и допускает ошибки при выполнении заданий, выполняет задание при подсказке преподавателя, затрудняется в формулировке выводов
2 «неудовлетворительно»	не способен правильно выполнить задание

7.3. Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности

Введение в проблематику построения систем анализа больших данных

Задания для решения кейса/комплект:

Кейс 1. Подготовка плана проекта по анализу данных

Требуется представить план проекта по анализу данных на заданную тему (например, прогнозирование оттока клиентов фирмы).

Работа состоит из следующих этапов:

- определение целей проекта и требований со стороны заказчика;
- определение цели анализа данных;
- разработка плана проекта;
- распределение обязанностей участникам проекта в зависимости от выполняемой роли.

Методы и технологии анализа больших данных

Лабораторные работы

1. Методы сбора открытых данных из сети интернет
2. Разведочный анализ и визуализация данных
3. Разработка предсказательной модели
4. Анализ поведения пользователей в сети интернет

Кейс 2. Разработка требований заказчика к информационно-аналитической системе

Цель: освоение базовых инструментов составления требований заказчика.

Требуется описать в 3-5 пользовательских историях и 1-2 сценариях использования любое приложение на ноутбуке или мобильном телефоне. Приветствуется добавление нового функционала.

Кейс 3. Проведение переговоров с подрядчиками и поставщиками оборудования и программного обеспечения

Требуется определить требования и составить план переговоров с подрядчиками и поставщиками оборудования и программного обеспечения

Обучение на больших данных

Лабораторные работы

1. Проектирование и обучение глубокой нейронной сети по распознаванию образов
2. Разработка нечетко-логического регулятора многомерного процесса

Пример задания лабораторной работы

Лабораторная работа №1 «Методы сбора открытых данных из сети интернет»

Цель: получить навыки применения различных методов и инструментов для сбора открытых данных из сети интернет.

Порядок выполнения:

1. Сбор данных с использованием API
Осуществить сбор данных с любого сайта, предоставляющего API-интерфейс.
Проанализировать состав данных, которые могут быть получены с помощью данного API-интерфейса.
2. Сбор данных без использования API
Осуществить сбор данных с любого сайта, не предоставляющего API-интерфейс, с помощью инструментов парсинга сайтов.
3. Проанализировать возможные варианты использования полученной информации.

Перечень вопросов к экзамену

1. Понятие и классификация больших данных. Размерность и объем данных. «Проклятие размерности».
2. Понятие и формирование озера данных Data Lake.
3. Методология исследования данных CRISP-DM. Задачи каждой фазы и возможные ошибки.
4. Методология исследования данных SEMMA. Отличия от CRISP-DM.
5. Компетенции и состав команды для работы с данными.
6. Правовые аспекты организации защиты персональных данных регламент защиты персональных данных (GDPR).
7. Задачи и методы извлечения знаний из больших данных.
8. Методы сбора данных из различных источников.
9. Методы предобработки данных.
10. Анализ социальных сетей.
11. Визуализация данных.
12. Назначение и характеристика компонент экосистемы Hadoop.
13. Парадигма MapReduce.
14. Apache Spark и его компоненты.
15. Облачные платформы для работы с большими данными.
16. Масштабирование и многоуровневое хранение данных: Парадигма NoSQL.
17. Особенности представления и обработки символьной и численной информации в нейронных сетях.
18. Методы и средства аннотирования данных для обучения нейронных сетей.
19. Современные алгоритмы и принципы обучения нейронных сетей, особенности обучения сетей различной структуры
20. Методы оценки качества нейросетевых моделей.
21. Валидация и интерпретируемость результатов работы нейросетевых моделей.
22. Системы нечеткого вывода: нечеткие высказывания, вывод в нечеткой логике, правила нечетких продукций.
23. Понятие нечеткой нейронной сети: определение и виды архитектур.
24. Обучение и алгоритмы оптимизации нечеткой нейронной сети.
25. Центры обработки данных: назначение, оборудование, уровни.

7.4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Фонды оценочных средств по дисциплине

Фонд оценочных средств позволяет оценить знания, умения и уровень приобретенных компетенций.

Фонд оценочных средств по дисциплине включает:

- вопросы к экзамену;
- комплект заданий к лабораторным работам.

Оценка качества освоения программы дисциплины включает текущий контроль успеваемости, промежуточную аттестацию, итоговую аттестацию.

Отчет по лабораторной работе

Содержание отчета:

1. Цель и задачи лабораторной работы.
2. Примеры запросов через API с описанием используемых параметров.
3. Примеры ответов по запросам и анализ полученной информации
4. Описание этапов парсинга сайта с указанием проводимых настроек
5. Анализ полученной информации.
6. Обосновать выбор модуля для ИИ.
7. Описать выявленные зависимости в данных.

Отчет по лабораторной работе представляется в электронном виде. Защита отчета проходит в форме доклада студента по выполненной работе и ответов на вопросы преподавателя. В случае, если оформление отчета и поведение студента во время защиты соответствуют указанным требованиям, студент получает максимальное количество баллов.

Основаниями для снижения количества баллов в диапазоне от max до min являются:

- небрежное выполнение,
- отсутствие выводов,
- нарушение сроков предоставления отчета.

Отчет не может быть принят и подлежит доработке в случае:

- серьезного несоответствия техническому заданию;
- отсутствия минимально необходимого количества тестовых примеров;
- некорректной работы программы и т.п.

Кейс/ситуационное задание

- Описание технологии применения кейса:
- задание выдается и выполняется во время лекционных занятий после изложения соответствующего теоретического материала;
- форма проведения - групповой анализ конкретной ситуации,
- время, отводимое на выполнение задания, - 30 минут.
- форма представления результатов — устный доклад и письменный отчет по результатам обсуждений.

Экзамен

Основаниями для снижения оценки являются:

- ошибки в объяснениях и комментариях при верно выполненном задании;
- неполный ответ;
- наличие мелких неточностей или незначительных искажений фактов.

В соответствии с балльно-рейтинговой системой БАРС по дисциплине отводится 100 баллов (50 баллов на семестровую часть: 40 баллов – текущие формы контроля и до 10 баллов – на бонусы; 50 баллов – на экзаменационную часть).

Текущий контроль осуществляется в ходе учебного процесса и консультирования студентов, по результатам выполнения соответствующих работ. Он предусматривает проверку готовности студентов к плановым занятиям, оценку качества и самостоятельности выполнения заданий на

практических занятиях, проверку правильности решения задач, выданных на самостоятельную проработку.

На экзамене осуществляется комплексная проверка знаний, навыков и умений студентов по материалу дисциплины на основании ответов на теоретические вопросы и решения практических задач.

Преподаватель, реализующий дисциплину, в зависимости от уровня подготовленности, обучающихся может использовать иные формы, методы контроля и оценочные средства, исходя из конкретной ситуации.

8. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

а) Основная литература:

1. Парфенов Ю.П. Постреляционные хранилища данных: учебное пособие для вузов / Ю.П. Парфенов; под научной редакцией Н.В. Папуловской. – Москва: Издательство Юрайт, 2021. – 121 с. – Текст : электронный // Образовательная платформа Юрайт [сайт]. – URL: <https://urait.ru/bcode/472624>.
2. Горбаченко В.И. Интеллектуальные системы: нечеткие системы и сети: учебное пособие для вузов / В.И. Горбаченко, Б.С. Ахметов, О.Ю. Кузнецова. – 2-е изд., испр. и доп. – Москва: Издательство Юрайт, 2021. – 105 с. – Текст : электронный // Образовательная платформа Юрайт [сайт]. – URL: <https://urait.ru/bcode/472491>.

б) Дополнительная литература:

3. Келлехер Дж. Наука о данных : Базовый курс / Дж. Келлехер, Б. Тирни. - Москва : Альпина Паблишер, 2020. - 222 с. - ISBN 978-5-9614-3170-4. - Текст : электронный // ЭБС «Консультант студента» : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785961431704.html>
4. Лесковец Ю. Анализ больших наборов данных / Лесковец Ю. , Раджараман А. , Джефффри Д. Ульман - Москва : ДМК Пресс, 2016. - 498 с. - ISBN 978-5-97060-190-7. - Текст: электронный // ЭБС «Консультант студента»: [сайт]. - URL: <https://www.studentlibrary.ru/book/ISBN9785970601907.html>
5. Орлов Г. М., Игнатъева О. А., Васин А. Г., Низомутдинов Б. А. Современные методы обработки и анализа данных. – СПб.: Университет ИТМО, 2021. – 147 с. – URL: – URL: books.ifmo.ru/file/pdf/2699.pdf

в) Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимый для освоения дисциплины (модуля)

1. Электронно-библиотечная система (ЭБС) ООО «Политехресурс» «Консультант студента». Многопрофильный образовательный ресурс «Консультант студента» является электронной библиотечной системой, предоставляющей доступ через сеть Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Каталог содержит более 15 000 наименований изданий. www.studentlibrary.ru.
2. Электронная библиотечная система издательства ЮРАЙТ, раздел «Легендарные книги». <https://urait.ru/>

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Учебные аудитории, библиотека АГУ, компьютерные классы, мультимедийные аудитории.

При необходимости рабочая программа дисциплины (модуля) может быть адаптирована для обеспечения образовательного процесса инвалидов и лиц с ограниченными возможностями здоровья, в том числе для обучения с применением дистанционных образовательных технологий. Для этого требуется заявление студента (его законного представителя) и заключение психолого-медико-педагогической комиссии (ПМПК).