

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
ОБРАБОТКА И АНАЛИЗ ДАННЫХ**

Составитель(-и)

**Михайлова Е.Г., к.ф.м.н., доцент, ИТМО
Выборнова О.Н., к.т.н, доцент каф. ИБ, АГУ
Коломина М.В., к.ф.м.н., доцент каф. ПМИ, 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

1

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1. Целями освоения дисциплины «Обработка и анализ данных» является получение компетенций для решения цифровых задач в профессиональной деятельности.

1.2. Задачи освоения дисциплины (модуля):

- изучение методов работы с данными, их области применения, достоинств и недостатков;
- сформировать культуру представления, описания, интерпретации и оценки выводов над данными.

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОПОП

2.1. Учебная дисциплина Б1.В.06 «Обработка и анализ 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:

- Теория вероятностей (уровень бакалавриата);
- Базы данных (уровень бакалавриата)

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

- Архитектура систем ИИ;
- Жизненный цикл цифровых продуктов на основе ИИ;
- Методология инженерии программных систем ИИ;
- Анализ и обучение на больших данных;
- Структурирование, разметка и обогащение данных;
- Искусственный интеллект для генерации изображений;
- Инфраструктуры больших данных;
- Интеллектуальный анализ данных;
- Параллельная и распределенная обработка данных;
- Глубокое обучение.

Также дисциплина «Обработка и анализ данных» поможет студентам при реализации задач производственной практики и написанию магистерской диссертации.

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-1. Способен адаптировать и применять на практике классические и новые научные принципы и методы исследований для решения задач в области создания и применения технологий и систем искусственного интеллекта.
- ПК-2. Способен использовать методы научных исследований и математического моделирования в области проектирования и управления системами искусственного интеллекта.
- ПК-5. Способен выбирать и применять методы инженерии знаний для создания систем, основанных на знаниях.

Таблица 1.
Декомпозиция результатов обучения

Код компетенции, индикатора	Планируемые результаты освоения дисциплины (модуля)		
	Знать (1)	Уметь (2)	Владеть (3)
ПК-1.1. Адаптирует известные научные	ПК-1.1.1 фундаментальные научные принципы и методы	ПК-1.1.2. адаптировать с целью практического при-	ПК-1.1.3 решать профессиональные задачи

принципы и методы исследований с целью их практического применения.	исследований	менения фундаментальные и новые научные принципы и методы исследований	на основе применения новых научных принципов и методов исследования
ПК-1.2. Решает профессиональные задачи на основе применения новых научных принципов и методов исследования.	ПК-1.2.1. особенности решения профессиональные задачи на основе применения новых научных принципов и методов исследования	ПК-1.2.2 разрабатывать, контролировать, оценивать и исследовать компоненты профессиональной деятельности; планировать самостоятельную деятельность в решении профессиональных задач	ПК-1.2.3. решать профессиональные задачи в области создания и применения технологий и систем искусственного интеллекта.
ПК-2. ПК-2.1. Применяет логические методы и приемы научного исследования, методологические принципы современной науки, направления, концепции, источники знания и приемы работы с ними, основные особенности научного метода познания, программно-целевые методы решения научных проблем в профессиональной деятельности.	ПК-2.1.1. логические методы и приемы научного исследования; методологические принципы современной науки, направления, концепции, источники знания и приемы работы с ними; основные особенности научного метода познания	ПК-2.1.2. применять логические методы и приемы научного исследования; методологические принципы современной науки, концепции, источники знания и приемы работы с ними; основные метода научного познания; программно-целевые методы решения научных проблем.	ПК-2.1.3. применения программно-целевых методов решения научных проблем в профессиональной деятельности.
ПК-5 ПК-5.1. Выбирает и применяет методы сбора и извлечения знаний.	ПК-5.1.1 методологические подходы к выбору и разработке методов получения знаний инженером по знаниям от экспертов; извлечения знаний из данных и текстов и применения соответствующих инструментальных средств.	ПК-5.1.2 методологические подходы к выбору и разработке методов получения знаний инженером по знаниям от экспертов; извлечения знаний из данных и текстов	ПК-5.1.3 навыками использования инструментов извлечения знаний из данных и текстов.
ПК-5 ПК-5.2. Выбирает и применяет методы структурирования знаний.	ПК-5.2.1 методологические подходы к выбору и применению методов структурирования знаний для предметных областей в виде ментальных карт, таксономий, деревьев целей и решений.	ПК-5.2.2 выбирать и применять методы структурирования знаний для построения концептуальных моделей знаний (онтологий знаний).	ПК-5.2.3 навыками использования инструментов структурирования знаний в виде ментальных карт, таксономий, деревьев целей и решений.
ПК-5 ПК-5.3. Выбирает и применяет методы представления знаний.	ПК-5.3.1 методологические подходы к выбору и применению методов представления знаний с помощью логических и продукционных методов, семантических сетей и фреймов, объектно-ориентированных методов.	ПК-5.3.2 выбирать и применять методы представления знаний для проектирования базы знаний для предметных областей.	ПК-5.3.3 навыками построения баз знаний на основе формальных языков представления знаний.
ПК-5 ПК-5.4. Выбирает и применяет методы обработки и распространения знаний.	ПК-5.4.1 методологические подходы к выбору и применению методов обработки и распространения знаний с помощью дедукции, индукции и абдукции, согласования экспертных оценок и нечеткого вывода.	ПК-5.4.2 выбирать и применять методы обработки и распространения знаний для разработки программных компонентов систем, основанных на знаниях, и приложений.	ПК-5.4.3 навыками разработки поисковых запросов к базам знаний, использование средств структурированного поиска.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Объем дисциплины (модуля) в зачетных единицах **2 з. е.** Всего 72 часа: 28 часов выделено на контактную работу обучающихся с преподавателем (лекции – 14, лабораторные работы – 14), 44 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 Форма промежуточной аттестации
				Л	ПЗ	ЛР	КР	СР	
1	Первичная обработка и хранение данных	1		7		7		22	Контрольная работа 1, Упражнение 11
2	Элементы статистического анализа данных			7		7		22	Контрольная работа 2, Упражнение 12
ИТОГО				14		14		44	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-1	ПК-2	ПК-5	общее количество компетенций
Первичная обработка и хранение данных	36	+	+	+	3
Элементы статистического анализа данных	36	+	+	+	3
Итого	72				

Краткое содержание дисциплины

Тема 1. Первичная обработка и хранение данных

Виды и источники данных. Загрузка и разделение данных. Объединение данных из разных источников. Очистка данных и заполнение пропусков. Контроль диапазонов. Инструменты для обработки данных. Визуализация данных. Формы представления количественных и качественных данных. Когнитивная визуализация данных. Первичная обработка данных. Сглаживание и нормировка данных. Преобразование данных. Обработка временных рядов. NoSQL базы данных. Большие данные. Хранение и доступ к данным. Виды баз данных. Реляционные СУБД.

Тема 2. Элементы статистического анализа данных

Основы теории вероятностей. Случайные события, понятия и действия над ними. Одномерные и многомерные случайные величины, их типы распределений. Интервальное оценивание и проверка гипотез, Точечное оценивание в параметрических и непараметрических случаях, сравнение оценок, эффективность. Эмпирическое распределение и описательная статистика. Действия со случайными величинами. Функции случайных величин. Характеристики случайных величин, характеристики зависимости. ЦПТ и ЗБЧ.

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 ма-

териала 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.

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.

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 раздела (темы)	Темы/вопросы, выносимые на самостоятельное изучение	Кол-во часов	Формы работы
1	Первичная обработка и хранение данных	22	Изучение теоретического материала. Подготовка к контрольной работе. Выполнение упражнений
2	Элементы статистического анализа данных	22	Изучение теоретического материала. Подготовка к контрольной работе. Выполнение упражнений

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

Выполнение индивидуального задания.

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.
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	Самостоятельная работа	Работа с ресурсами Internet.

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 «Обработка и анализ 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 ре-

зультатов 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.	Первичная обработка и хранение данных	ПК-1, ПК-2, ПК-5	Контрольная работа 1, Упражнение 1
2.	Элементы статистического анализа данных	ПК-1, ПК-2, ПК-5	Контрольная работа 2, Упражнение 2

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

Тема 1. Первичная обработка и хранение данных

Упражнение 1.

После каждой лекции студентам выдаются индивидуальные упражнения. Задание выполняется студентом при помощи сети Интернет и инструментов, рассматриваемых в лекции.

Пример задания: Есть исходные данные для анализа следующего вида:

Имя	Возраст	Пол	Образование	Семейное положение
Иннокентий	35	м	высшее	женат
Семен	34	м	высшее	женат
Анфиса	45	ж	среднее	замужем
Светлана	15	ж	начальное	не замужем
Инна	19	ж	среднее	не женат
Михаил	27	м	высшее	разведен
Александр	25	м	высшее	разведен
Екатерина	27	ж	среднее	замужем

Каким шкалам соответствуют значения в каждой из колонок? Ответы выбирать из возможных значений: номинальная, порядковая, интервальная, относительная, дихотомическая

Контрольная работа 1.

Студенты получают доступ к таблицам тестовым базам данных (Oracle и MongoDB):

STUDENT_PROFILE(student_number, student_fio, student_birthday), STUDENT_MARKS(student_number, subject_code, subject_mark, subject_mark_date) STUDENT_PUBLICATIONS(student_number, publication_id, publication_name, publication_issue_date)

Требуется:

1. Выполнить контроль диапазонов отметок (**subject_mark** должен быть в интервале от 1 до 5). Исключить неподходящие записи.
2. Выполнить контроль публикаций студента и дат полученных отметок. Исключить публикации и отметки с сомнительной датой (опубликованные/полученные в возрасте менее 5 лет или с еще не наступившей датой).
3. Объединить данные из нескольких таблиц в одну таблицу **STUDENT_TOTAL** (student_number, student_fio, mark_5_count, mark_4_count, publication_count), где **mark_5_count** – количество отметок 5, **mark_4_count** - количество отметок 4, **publication_count** – количество публикаций.
4. Нормировать значения полей **mark_5_count**, **mark_4_count** и **publication_count** заданным способом нормировки (с точностью до двух знаков после десятичного разделителя).
5. Из базы MongoDB получить документ по заданному значению параметра.
6. Вывести в качестве результата сумму нормированных значений **mark_5_count**, **mark_4_count** и **publication_count**.

Тема 2. Элементы статистического анализа данных

Упражнение 2.

Задание выполняется студентом при помощи сети Интернет и инструментов, рассматриваемых в лекции.

- найдите выборочное среднее; найдите оценку среднеквадратического отклонения;
- найдите значение модуля функции отклонения;
- ответьте на вопрос, верно ли заявления (справедлива ли выдвинутая гипотеза).

Контрольная работа 2.

Упражнение выполняется студентом при помощи сети Интернет и инструментов, рассматриваемых в лекции.

В Базе Данных (инструкция по доступу) есть таблица pulsar_stars, в которой содержатся сведения о звездах, полученные в ходе исследования вселенной (High Time Resolution Universe

Survey) с целью определения одного из типа нейтронных звезд – пульсаров. Поле TARGET таблицы pulsar_stars является откликом, все остальные поля – предикторы.

Вам необходимо получить выборку из таблицы pulsar_stars с помощью запросов на основании следующих критериев:

- Все строки таблицы, где TARGET = 0 и MIP ∈ [87.5859375, 88.484375];
- Все строки таблицы, где TARGET = 1 и MIP ∈ [52.9296875, 58.7890625].

Укажите число строк в полученной выборке.

Определите выборочное среднее для столбца MIP.

Выполните линейную нормировку всех значений предикторов полученной выборки. Определите выборочное среднее для столбца MIP после нормировки.

Обучите модель логистической регрессии, используя полученную после нормировки выборку в качестве тренировочного набора данных. Выполните классификацию новой звезды с параметрами: [0.833, 0.092, 0.443, 0.092, 0.112, 0.86, 0.742, 0.299]

Введите вероятность отнесения звезды к классу пульсар.

Выполните классификацию новой звезды, с помощью метода k-ближайших соседей, используя нормализованные данные выборки.

Введите расстояние от новой звезды до ближайшего соседа, используя евклидову метрику.

Экзамен

Контрольная работа выполняется студентом в день экзамена и включает в себя задания, аналогичные заданиям из текущего контроля успеваемости. Время выполнения ограничено 60-90 минутами. Варианты контрольной работы генерируются автоматически и представляют собой задачи как с закрытым, так и открытым типом вопросов. Контрольная работа включает в себя некоторое количество (5-20) заданий. Задания даются по теме «Первичная обработка и хранение данных» и из темы «Элементы статистической обработки данных»

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.

Итоговая оценка успеваемости студентов по дисциплине производится согласно Положению о балльно-рейтинговой системе оценки учебных достижений студентов, утверждено приказом ректора от 13.01.2014 г. № 08-01-01/08.

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. Анализ данных: учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под ред. В. С. Мхитаряна. – М.: Издательство Юрайт, 2018. – 490 с. – (Серия : Бакалавр. Акаде-

- мический курс). – ISBN 978-5-534-00616-2. – Режим доступа: www.biblio-online.ru/book/CC38E97A-CCE5-4470-90F1-3B6D35ACC0B4.
2. Миркин Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. М.: Издательство Юрайт, 2018. – 174 с. – (Серия: Авторский учебник). – ISBN 978-5-9916-5009-0. – Режим доступа: www.biblio-online.ru/book/46A41F93-BC46-401C-A30E-27C0FB60B9DE.
 3. Обработка и анализ данных / Университет ИТМО – Открытое образование. – Режим доступа: <https://openedu.ru/>
 4. Элементы статистической обработки данных / Университет ИТМО – Открытое образование. – Режим доступа: <https://openedu.ru/>

б) Дополнительная литература:

1. Неделько В. М. Основы статистических методов машинного обучения: учебное пособие / В. М. Неделько. – Новосибирск: Новосибирский государственный технический университет, 2010. – 72 с. – ISBN 978-5-7782-1385-2. – Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. – URL: <https://www.iprbookshop.ru/45418.html>

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

1. Электронная библиотечная система издательства ЮРАЙТ, раздел «Легендарные книги». www.biblio-online.ru
2. Электронная библиотечная система IPRbooks. www.iprbookshop.ru
3. Обработка и анализ данных / Университет ИТМО – Открытое образование. – Режим доступа: <https://openedu.ru/>
4. Элементы статистической обработки данных / Университет ИТМО – Открытое образование. – Режим доступа: <https://openedu.ru/>
5. Введение в машинное обучение / Университет ИТМО – Открытое образование. – Режим доступа: <https://openedu.ru/>

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Учебные аудитории, библиотеки 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).