

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Астраханский государственный университет имени В. Н. Татищева»
(Астраханский государственный университет им. В. Н. Татищева)

СОГЛАСОВАНО
Руководитель ОПОП

_____ М.В. Коломина

«8» сентября 2022 г.

УТВЕРЖДАЮ
Зав. кафедрой ПМИ

_____ М.В. Коломина

«8» сентября 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Анализ данных»

Составители	Станкевич А. С., к.т.н., доцент ФИТиП, ИТМО Корнеев Г. А., к.т.н., доцент ФИТиП, ИТМО
Направление подготовки / специальность	01.03.02 Прикладная математика и информатика
Направленность (профиль) ОПОП	Программирование и искусственный интеллект
Квалификация (степень)	бакалавр
Форма обучения	очная
Год приёма	2023
Курс	3
Семестр(ы)	6

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1. Целями освоения дисциплины «Анализ данных» является освоение практических аспектов технологий, связанных с хранением, обработкой, подходами к анализу больших объемов данных.

1.2. Задачи освоения дисциплины:

- изучение вопросов, связанных с хранением и первичной обработкой данных;
- изучение вопросов статистического и машинного обучения;
- формирование практических навыков решения прикладных задач.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

2.1. Учебная дисциплина «Анализ данных» относится к части, формируемой участниками образовательных отношений и осваивается в 6 семестре.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

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

а) общепрофессиональных (ОПК);

ОПК-3. Способен формулировать, строить и применять математические модели для управления достижением планируемых результатов процессов и объектов профессиональной деятельности на базе знаний математики, программирования и унифицированных пакетов программ.

б) профессиональных (ПК).

ПК-8. Способность понимать, совершенствовать и применять современный математический аппарат.

ПК-14. Способность определять эффективный способ решения прикладных задач с применением информационных технологий и программной инженерии, разрабатывать и внедрять соответствующие программные решения.

ПК-18. Способен разрабатывать и применять методы машинного обучения для решения задач.

ПК-21. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта.

ПК-23. Способен принимать участие в управлении проектами по созданию и развитию технологий и систем искусственного интеллекта на стадиях их жизненного цикла.

Таблица 1 – Декомпозиция результатов обучения

Код и наименование компетенции	Планируемые результаты обучения по дисциплине (модулю)		
	Знать (1)	Уметь (2)	Владеть (3)
ОПК-3.1 Выявляет и формулирует целевые характеристики объекта моделирования в профессиональной деятельности ОПК-3.2 Определяет методы описания объектов и соответствующие им модели в профессиональной деятельности ОПК-3.3 Строит модели объектов и процессов профессиональной деятельности	математические модели, моделирование, методы описания объектов	выявлять и формулировать целевые характеристики описания объекта моделирования, определять методы описания объектов и соответствующие им модели, строить модели объектов и процессов, апробировать и реализовывать математические модели в программной среде, осуществлять их корректировку, применять модели объектов и процессов, оценивать достижение целевых характеристик и показателей,	навыками описания, построения, применения моделей объектов, оценки и интерпретации результатов моделирования процессов и объектов

<p>сти на базе знаний математики, программирования и унифицированных пакетов программ</p> <p>ОПК-3.4 Апробирует и реализует математические модели в программной среде и осуществляет их корректировку (при необходимости)</p> <p>ОПК-3.5 Применяет модели объектов и процессов, оценивает достижение целевых характеристик и показателей в профессиональной сфере</p> <p>ОПК-3.6 Интерпретирует и представляет результаты моделирования процессов и объектов профессиональной деятельности</p>		<p>интерпретировать и представлять результаты моделирования процессов и объектов в профессиональной деятельности</p>	
<p>ПК-8.1. Владение методами интегрального и дифференциального исчисления одной и нескольких переменных</p> <p>ПК-8.2. Владение методами теории линейных пространств и операторов</p> <p>ПК-8.3. Владение методами функционального анализа для решения сложных задач информатики</p>	<p>современный математический аппарат</p>	<p>владеть методами функционального анализа для решения сложных задач информатики</p>	<p>навыками применения современного математического аппарата</p>
<p>ПК-14.1. Способность проектировать и реализовывать программные решения с применением методов функционального, автоматного и эволюционного программирования</p>	<p>методы функционального, автоматного и эволюционного программирования</p>	<p>проектировать и реализовывать мобильные и web-приложения</p>	<p>навыками определения эффективного способа решения прикладных задач с применением информационных технологий и программной инженерии, разрабатывать и внедрять соответствующие программные решения</p>
<p>ПК-18.</p> <p>ПК-18.1. Проводит анализ требований и определяет необходимые классы задач машинного обучения</p>	<p>принципы и методы машинного обучения, типы и классы задач машинного обучения, методологию ML Ops; статистические методы анализа данных</p>	<p>сопоставить задачам предметной области классы задач машинного обучения</p>	<p>методами машинного обучения и статистическими методами анализа данных</p>
<p>ПК-18.</p> <p>ПК-18.2. Принимает участие в оценке, выборе и при необходимости разработке методов машинного обучения</p>	<p>классические методы и алгоритмы машинного обучения: предиктивные - обучение с учителем, обучение без учителя</p>	<p>проводить сравнительный анализ и осуществлять выбор, настройку при необходимости разработку методов и алгоритмов для решения задач машинного обучения</p>	<p>методами обучения с учителем, обучение без учителя</p>

<p>ПК-21.1. Осуществляет поиск данных в открытых источниках. Специализированных библиотеках и репозиториях</p> <p>ПК-21.2. Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения</p>	<p>виды представления данных, методы поиска и парсинга данных</p> <p>уровни представления данных (ODS, DDL, семантический слой, модель данных)</p> <p>основные инструменты, библиотеки и технологии data Science</p> <p>методы редукции размерности элементов набора данных и их предварительной статистической обработки разметки структурированных и неструктурированных данных</p>	<p>отделять достоверные источники данных от сомнительных, осуществлять критический отбор данных, проверять их целостность и непротиворечивость</p> <p>использовать инструменты и библиотеки для data science для поиска данных в открытых источниках, специализированных библиотеках и репозиториях</p> <p>выявлять и исключить из массива данных ошибочные данные и выбросы</p> <p>осуществлять разметку структурированных и неструктурированных данных</p> <p>использовать инструменты, библиотеки и технологии Data Science для подготовки и разметки структурированных и неструктурированных данных для машинного обучения</p> <p>использовать методы и технологии массово-параллельной обработки и анализа данных</p>	<p>методами поиска и парсинга данных</p> <p>методами редукции размерности элементов набора данных и их предварительной статистической обработки разметки структурированных и неструктурированных данных,</p>
<p>ПК-23.1 Использует основы управления проектами по созданию и развитию технологий и систем искусственного интеллекта на стадиях их жизненного цикла</p>	<p>основы управления проектами по созданию и развитию технологий и систем искусственного интеллекта на стадиях их жизненного цикла</p>	<p>управлять проектами по созданию и развитию технологий и систем искусственного интеллекта на стадиях их жизненного цикла</p>	<p>основами управления проектами по созданию и развитию технологий и систем искусственного интеллекта на стадиях их жизненного цикла</p>

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Объём дисциплины составляет 4 зачётные единицы, в том числе 72 часа, выделенных на контактную работу обучающихся с преподавателем (из них 36 часов – лекции, 36 часов – лабораторные работы), и 72 часа – на самостоятельную работу обучающихся.

Таблица 2 – Структура и содержание дисциплины

Раздел, тема дисциплины (модуля)	Семестр	Контактная работа (в часах)			Самост. работа		Форма текущего контроля успеваемости, форма промежуточной аттестации
		Л	ПЗ	ЛР	КР	СР	
Хранение и первичная обработка данных	6	18	-	18	-	36	лабораторная работа контрольная работа упражнения
Статистическое и машинное обучение	6	18	-	18	-	36	лабораторная работа контрольная работа упражнения
Итого		36	-	36	-	72	Экзамен

Краткое содержание каждой темы дисциплины (модуля)

Раздел 1. Хранение и первичная обработка данных

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

Раздел 2. Статистическое и машинное обучение

Элементы математической статистики: точечные и интервальные оценки, проверка гипотез. Методы оценивания некоторых характеристик статистических связей, Регрессии и классификация: линейная, многомерная, полиномиальная, логистическая, наивный Байесовский классификатор, деревья решений и метод ближайших соседей, Метод опорных векторов и нейросети. Задачи машинного обучения (обучение с учителем, без учителя), Композиции: бэггинг, случайный лес, градиентный бустинг.

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРЕПОДАВАНИЮ И ОСВОЕНИЮ ДИСЦИПЛИНЫ

5.1. Указания для преподавателей по организации и проведению учебных занятий по дисциплине (модулю)

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 нарастание сложности 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).

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 раздела (темы)	Темы/вопросы, выносимые на самостоятельное изучение	Кол-во часов	Форма работы
Раздел 1	Хранение и первичная обработка данных	36	Изучение теоретического материала. Подготовка к лабораторным работам. Решение заданий
Раздел 2	Статистическое и машинное обучение	36	Изучение теоретического материала. Подготовка к лабораторным работам. Решение заданий

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

Экзамен проводится в письменной форме в виде контрольной работы.

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 разделу 1 «Хранение и первичная обработка данных» и разделу 2 – «Статистическое и машинное обучение». Максимальное количество баллов за экзамен – 20.

6. ОБРАЗОВАТЕЛЬНЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

При реализации различных видов учебной работы по дисциплине «Анализ данных» 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.
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, 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).

Перечень информационных справочных систем:

1. Электронная библиотека «Астраханский государственный университет» собственной генерации на платформе ЭБС «Электронный Читальный зал – БиблиоТех». <https://biblio.asu.edu.ru>
2. Электронный каталог Научной библиотеки АГУ на базе MARK SQL НПО «Информ-систем».
3. <https://library.asu.edu.ru>
4. Электронно-библиотечная система (ЭБС) ООО «Политехресурс» «Консультант студента». www.studentlibrary.ru
5. Электронная библиотечная система издательства ЮРАЙТ, раздел «Легендарные книги». www.biblio-online.ru, <https://urait.ru/>
6. Электронная библиотечная система IPRbooks. www.iprbookshop.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 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 и прохождением практик, а в процессе освоения дисциплины – последовательным достижением результатов освоения содержательно связанных между собой разделов, тем.

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

Таблица 5 – 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

Шкала оценивания	Критерии оценивания
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 6 – 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, испытывает затруднения и допускает ошибки при выполнении заданий, выполняет задание по подсказке преподавателя, затрудняется в формулировке выводов
2 «неудовлетворительно»	не способен правильно выполнить задания

7.3. Контрольные задания и иные материалы, необходимые для оценки результатов обучения по дисциплине (модулю)

Типовые контрольные задания, необходимые для оценки достижения запланированных результатов обучения приведены в таблице планирования результатов обучения по дисциплине (БаРС) (Приложение 1)*.

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств:

Контрольная работа 1

Описание технологии проведения контрольной работы:

Контрольная работа выполняется в форме тестов. После каждой лекции студентам выдаются 1-3 тестовых задания. Срок выполнения теста – не позднее срока освоения дисциплины.

Примеры контрольных заданий для блока 1:

1. Есть исходные данные для анализа следующего вида:

<i>Имя</i>	<i>Возраст</i>	<i>Пол</i>	<i>Образование</i>	<i>Семейное положение</i>
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 (3 балла)

0. Для заданного набора данных (таблица выше) определить распределение в процентах параметра «Образование» (4 балл)

1. Исходные данные:

	2014	2015	2016	2017	2018
Мужчины	13456	11344	12444	14333	12555
Женщины	11333	12323	11212	12121	13121
Пенсионеры	132	153	153	211	242
Студенты	232	345	364	374	435
Пожарники	22	35	38	58	45

Для заданной таблицы данных указать, какие строки можно агрегировать. Ответы выбираются из возможных комбинаций строк (3 балла)

Шкала оценивания и критерии оценки за контрольные работы блока 1:

Минимальное количество баллов — 0 баллов

Максимальное количество баллов — 10 баллов

Оценка снижается при отсутствии развёрнутого решения, наличия доказательных и вычислительных ошибок.

Лабораторная работа 1

Лабораторная работа выполняется в рамках каждого раздела курса с целью усвоения прослушанного студентом теоретического материала. Лабораторная работа реализуется в форме тестовых вопросов внутри лекции.

Объем выполненной работы: каждая лабораторная работа содержит 3-5 вопросов.

Срок сдачи работы указывается преподавателем. Студент получает вопросы и предлагаемые ответы, из которых необходимо выбрать правильные. На выполнение работы дается одна попытка.

Примеры вопросов для лабораторной работы «Хранение и первичная обработка данных»:

1. Укажите шкалы, которые являются дихотомическими:
 - Мнение ('согласен' / 'не могу ответить' / 'не согласен')
 - Пол ('мужской' / 'женский')
 - Возраст ('младше 10 лет' / 'от 10 до 20 лет' / 'от 20 до 30 лет' / 'от 30 до 40 лет' / 'старше 40 лет')

0. Укажите свойства ключа таблицы:
- Может принимать незаданные значения
 - Может повторяться в разных строках таблицы
 - Должен быть только числовым значением
 - Должен принимать уникальные значения для каждой строки таблицы
 - Может состоять из одного поля, а может из нескольких полей
0. Выберите верные высказывания:
- Мода всегда больше медианы
 - Мода всегда равна медиане
 - Средневзвешенное значение всегда равно моде
 - Средневзвешенное значение всегда равно медиане
 - Все предыдущие утверждения неверны

Шкала оценивания и критерии оценки (на примере одной лабораторной работы):

Требования	Максимальное количество баллов
1. Логика решения	5
2. Верно произведенные расчеты	5
Итого:	10

Упражнение по блоку 1

После всех лекций блока 1 студентам выдается одно упражнение (1-6 заданий).

Срок выполнения упражнений – не позднее срока освоения дисциплины. Для выполнения задания дается две попытки. За правильное решение со второй попытки количество баллов уменьшается (обычно в два раза). Итоговый балл за упражнения подсчитывается по максимуму из двух попыток целым числом с округлением в большую сторону.

Пример упражнений блока 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 балла)
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). (2 балла)
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. (5 баллов)

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 баллов)
5. Из базы MongoDB получить документ по заданному значению параметра. (4 балла)
6. Вывести в качестве результата сумму нормированных значений **mark_5_count**, **mark_4_count** и **publication_count**. (4 балла)

Критерии оценивания:

Требования	Максимальное количество баллов
1. Логика решения	10
2. Верно произведенные расчеты	10
Итого:	20

Лабораторная работа 2

Лабораторная работа выполняется в рамках каждого раздела курса с целью усвоения прослушанного студентом теоретического материала. Лабораторная работа реализуется в форме тестовых вопросов внутри лекции.

Объем выполненной работы: каждая лабораторная работа содержит 3-5 вопросов.

Срок сдачи работы указывается преподавателем. Студент получает вопросы и предлагаемые ответы, из которых необходимо выбрать правильные. На выполнение работы дается одна попытка.

Примеры вопросов для лабораторной работы «Статистическое и машинное обучение»:

1. Дискретная случайная величина X принимает значения $\{-1, 0, 1\}$ с вероятностями соответственно $\{0.1, 0.6, 0.3\}$. Ее математическое ожидание равно:
 - 2.2
 - 2.4
 - 0.2
 - 1.2
 - 0.6
2. Логистическая функция задается выражением:
 - $\sigma(x) = \frac{1}{1 + e^{-x}}$
 - $f(s) = \frac{s}{|s| + \alpha}$
 - $f(s) = \frac{s}{|s| + \alpha}$
2. Какие из этих задач являются типичными для машинного обучения без учителя?
 - Оценка вероятности выхода автомобиля из строя
 - Предсказание вероятности покупки автомобиля
 - Группировка изображений по визуальным признакам
 - Рекомендация видеороликов
2. Случайная величина Y задается биномиальным распределением с параметрами $n=10$ и $p=1/5$. Ее дисперсия равна:
 - 5

- 1/2
 - 4/5
 - 2
 - 8/5
2. Бэггинг – это метод, позволяющий снизить:
- ошибку разброса
 - ошибку смещения
 - ошибку смещения и ошибку разброса
 - не подходит ни один вариант из приведенных

Шкала оценивания и критерии оценки (на примере одной лабораторной работы):

Требования	Максимальное количество баллов
1. Логика решения	5
2. Верно произведенные расчеты	5
Итого:	10

Контрольная работа 2

Описание технологии проведения контрольной работы:

Контрольная работа выполняется в форме тестов

После каждой лекции студентам выдаются 1-3 тестовых задания.

Срок выполнения теста – не позднее срока освоения дисциплины.

Примеры контрольных заданий для блока 2:

На основе анализа некоторого числа писем электронной почты сформирована таблица 1, содержащая информацию о классификации писем на группы «спам» и «не спам», а также суммарное число слов входящих в эти группы. В таблице 2 представлены данные, по уникальным словам, и числу их вхождений в указанные группы.

Таблица 1.

	SPAM	HAM
Emails	19	7
Words	114	42

Таблица 2.

	SPAM	HAM
Credit	21	2
Online	5	2
Bill	0	3

Free	11	11
Cash	3	4
Bonus	23	8
Remove	3	9
Money	12	0
Offer	23	2
Coupon	13	1

Ваша задача найти оценку апостериорного максимума и вероятность для определения наиболее вероятного класса, к которому будет отнесено письмо, содержащее текст:

«Online Credit Offer Access Cash Free Membership»

Правильный ответ может отклоняться от эталонных не более, чем на 2%.

Шкала оценивания и критерии оценки за контрольные работы блока 2:

Минимальное количество баллов — 0 баллов

Максимальное количество баллов — 10 баллов

Каждая правильно решённая задача оценивается в 5 баллов. Оценка снижается при отсутствии развёрнутого решения, наличия доказательных и вычислительных ошибок.

Упражнение по блоку 2

После всех лекций блока 2 студентам выдается одно упражнение (1-6 заданий).

Срок выполнения упражнений – не позднее срока освоения дисциплины. Для выполнения задания дается две попытки. За правильное решение со второй попытки количество баллов уменьшается (обычно в два раза). Итоговый балл за упражнения подсчитывается по максимуму из двух попыток целым числом с округлением в большую сторону.

Пример упражнений блока 2:

Для сгенерированного датасета выполнить указанные действия и расчеты.

1. Разделить выборку на тренировочную и тестовую (90% и 10%). Обучить классификатор на основе метода логистической регрессии. (5 баллов)
2. Оценить качество классификации на тренировочной/тестовой выборке. Сделать 10-fold кросс-валидацию и оценить дисперсию. (5 баллов)
3. Построить графики зависимости точности на тренировочном/тестовом наборе для n-fold кросс-валидации для разных значений $n = 2, 5, 8, 10$. (5 баллов)
4. Сделать вывод о сравнительном качестве предсказаний при разных способах оценки точности. (5 баллов)

Критерии оценивания:

Требования	Максимальное количество баллов
1. Логика решения	8
2. Верно произведенные расчеты	12
Итого:	20

7.4. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине (модулю)

ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств:

Экзамен

Экзамен проводится в письменной форме в виде контрольной работы.

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 разделу 1 «Хранение и первичная обработка данных» и разделу 2 – «Статистическое и машинное обучение». Максимальное количество баллов за экзамен — 20.

Раздел 1. Хранение и первичная обработка данных

1. К какому из приведенных типов шкал относится шкала Бофорта? Номинальная, , Интервальная, Относительная.
2. Какой оператор языка SQL позволяет добавить (вставить) строки в таблицы: CHANGE INTO, ALTER INTO, MODIFY INTO, ADD INTO, INSERT INTO

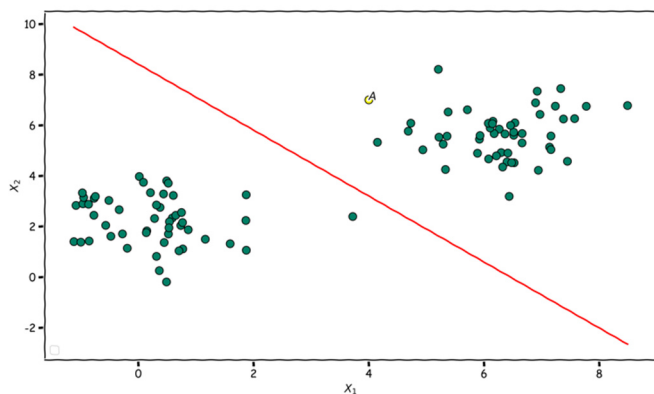
Раздел 2. Статистическое и машинное обучение Элементы статистической обработки данных

1. Функция распределения случайной величины ξ задана соотношением

$$F_{\xi}(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-3x}, & x > 0 \end{cases}$$

Какое распределение имеет данная случайная величина?

1. Exp_{-3}
 2. Exp_3
 3. $Exp_{1/3}$
 4. $Exp_{-1/3}$
0. Дана выборка $(1, 2, 1, 4, 2)$.
1. Найдите выборочное среднее.
 2. Найдите выборочную медиану.
0. Пусть $\xi \sim Exp_{\lambda}$. Введите выражение для оценки λ , используя выборочное среднее.
0. Уравнение линейной регрессии задается соотношением $3 + 4X_1 - 2X_2 = 0$. Предскажите отклик для тестового объекта — (6, 6).
0. Модель логистической регрессии обучена на определение класса +.
- $\Psi = -8.872 + 1.374 \cdot X_1 + 1.055 \cdot X_2$. На рисунке отображены точки, на которых обучена модель, но вам неизвестно какие из них относятся к классу + и -.



Определите, в каком направлении расположен нормальный вектор гиперплоскости. Учитывая, что тестовая точка А с координатами (4,7) относится к классу +, верно ли она классифицирована моделью?

Итоговая оценка по дисциплине рассчитывается на основании результатов (набранного числа баллов) каждого из разделов и промежуточной аттестации.

Оценка	Минимальное количество баллов	Максимальное количество баллов
«5» (отлично)	90	100
«4» (хорошо)	74	90
«3» (удовлетворительно)	60	74
«2» (неудовлетворительно)	0	60

Знания, умения и навыки обучающихся при промежуточной аттестации **в форме экзамена** определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

«Отлично» – обучающийся глубоко и прочно усвоил весь программный материал, исчерпывающе, последовательно, грамотно и логически стройно его излагает, не затрудняется с ответом при видоизменении задания, свободно справляется с задачами и практическими заданиями, правильно обосновывает принятые решения, умеет самостоятельно обобщать и излагать материал, не допуская ошибок.

«Хорошо» – обучающийся твердо знает программный материал, грамотно и по существу излагает его, не допускает существенных неточностей в ответе на вопрос, может правильно применять теоретические положения и владеет необходимыми умениями и навыками при выполнении практических заданий.

«Удовлетворительно» – обучающийся усвоил только основной материал, но не знает отдельных деталей, допускает неточности, недостаточно правильные формулировки, нарушает последовательность в изложении программного материала и испытывает затруднения в выполнении практических заданий.

«Неудовлетворительно» – обучающийся не знает значительной части программного материала, допускает существенные ошибки, с большими затруднениями выполняет практические задания, задачи.

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

8.1. Основная литература

1. Hopcroft J. E., Motwani R., Ullman J. D. Introduction to Automata Theory, Languages, and Computation (3rd Edition). — Addison-Wesley, Boston, MA, USA, 2006. — 750 с.
2. Шень А. Программирование: теоремы и задачи. — М.: МЦНМО, 2014. — 296 с.
3. Шень А., Верещагин Н. Языки и исчисления. — М.: МЦНМО, 2012. — 240 с.
4. Верещагин, Н. К. Колмогоровская сложность и алгоритмическая случайность [Электронный ресурс] / Н. К. Верещагин, В. А. Успенский, А. Шень. — Электрон. дан. — СПб: Лань, 2013. — 575 с. — Режим доступа: <https://e.lanbook.com/book/56395> — Загл. с экрана.

8.2. Учебно-методическое обеспечение для самостоятельной работы обучающихся:

1. Кривцова, И. Е. Основы дискретной математики. Часть 1. Учебное пособие [Электронный ресурс] / И. Е. Кривцова, И. С. Лебедев, А. В. Настека. — Электрон. дан. — СПб: ИТМО, 2016. — 92 с. — Режим доступа: http://books.ifmo.ru/book/1869/osnovy_diskretnoy_matematiki_chast_1_uchebnoe_posobie.htm — Загл. с экрана.

8.3. Дополнительная литература

1. Вики-конспекты. — http://neerc.ifmo.ru/wiki/index.php?title=Заглавная_страница

8.4. Перечень ресурсов информационно-телекоммуникационной сети “Интернет”, необходимый для освоения дисциплины

1. Электронный каталог Научной библиотеки АГУ на базе MARK SQL НПО «Информ-систем»: <https://library.asu.edu.ru>
2. Корпоративный проект Ассоциации региональных библиотечных консорциумов (АРБИКОН) «Межрегиональная аналитическая роспись статей» (МАРС): <http://mars.arbicon.ru>
3. Единое окно доступа к образовательным ресурсам <http://window.edu.ru>

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Для проведения лекционных занятий используется аудитория, оборудованная современной презентационной техникой (проектор, экран, ноутбук).

Для выполнения лабораторных работ используются компьютерные классы с установленным в них необходимым программным обеспечением.

При необходимости рабочая программа дисциплины 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).