

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
_____ Р.Ю. Демина

«06» июня 2024 г.

УТВЕРЖДАЮ
И.о. заведующего кафедрой ИБ

_____ Т.Г. Гурская

от «06» июня 2024 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКИЕ МЕТОДЫ В АНАЛИЗЕ
ДАННЫХ**

Составитель(-и)	Мартьянова А.Е., к.т.н., доцент кафедры информационной безопасности
Направление подготовки	10.03.01 ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ
Направленность (профиль) ОПОП	«Организация и технология защиты информации» (в сфере информационных и коммуникационных технологий)
Квалификация (степень)	бакалавр
Форма обучения	очно-заочная
Год приема	2023
Курс	2
Семестр	3

Астрахань, 2024

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1. Целью освоения дисциплины «Вероятностно-статистические методы в анализе данных» является формирование у студентов современных теоретических знаний о вероятностных и статистических закономерностях, практических навыков в решении и исследовании прикладных задач теоретико-вероятностного и статистического характера, выработка у студентов теоретико-вероятностной интуиции, необходимой при решении разнообразных задач, возникающих в профессиональной деятельности.

Изучение дисциплины «Вероятностно-статистические методы в анализе данных» способствует развитию у студентов склонности и способности к творческому мышлению, выработке системного подхода к исследуемым явлениям.

В ходе изучения дисциплины предусматривается проведение лекционных и лабораторных занятий, выполнение практических работ.

В лекциях излагается содержание тем программы с учётом требований, установленных для специалиста в квалификационной характеристике. Основное внимание уделяется наиболее сложным вопросам курса.

Лабораторные занятия проводятся в учебных группах с целью закрепления теоретических основ, излагаемых в лекционном курсе, и получения практических навыков в применении методов теории вероятностей и математической статистики в практической инженерной деятельности.

Бакалавр, изучив дисциплину «Вероятностно-статистические методы в анализе данных», должен решать следующие профессиональные задачи в соответствии с видами профессиональной деятельности:

проектно-технологическая деятельность:

- проведение проектных расчетов элементов систем обеспечения информационной безопасности,
- сбор и анализ исходных данных для проектирования систем защиты информации, определение требований, сравнительный анализ подсистем по показателям информационной безопасности,

экспериментально-исследовательская деятельность:

- проведение экспериментов по заданной методике, обработка и анализ результатов.

1.2. Задачи освоения дисциплины:

- 1) приобретение студентами необходимых знаний основ вероятностно-статистических методов в анализе данных;
- 2) овладение навыками исследования случайных величин, вычисления их основных характеристик, статистического анализа выборок;
- 3) приобретение знаний и навыков моделирования случайных событий, обработки статистических данных, точечного и интервального оценивания параметров распределений, проверки статистических гипотез;
- 4) формирование умения интерпретировать результаты вероятностных и статистических исследований и применять их при решении практических задач;
- 5) применение полученных знаний, навыков и умений в последующей профессиональной деятельности.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

2.1. Учебная дисциплина Б1.Б.10 «Вероятностно-статистические методы в анализе данных» относится к части, формируемая участниками образовательных отношений и осваивается в 3 семестре.

Изучение курса «Вероятностно-статистические методы в анализе данных» рассчитано на один семестр (3 семестр) и предусматривает сдачу студентами зачета на основе рейтинго-балльной системы оценивания. Общая трудоемкость дисциплины – 3 кредита (3 ЗЕ) 108 часов в 3-ем семестре.

Бакалавр, изучив дисциплину «Вероятностно-статистические методы в анализе данных», может быть готов к следующему виду профессиональной деятельности:

- эксплуатационная;
- проектно-технологическая;
- организационно-управленческая.

Бакалавр, изучив дисциплину «Вероятностно-статистические методы в анализе данных», должен решать следующие профессиональные задачи в соответствии с видами профессиональной деятельности:

- применение полученных знаний в научно-исследовательской и практической работе.
- определение возможных каналов утечки информации;
- осуществление противодействия несанкционированному информационному воздействию на вычислительные сети;
- обеспечение грамотного выбора программных и программно-аппаратных средств защиты информации;

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

- Математика.
- Информатика.

В результате освоения этих дисциплин, студент должен: знать:

- основные понятия и методы математического анализа;
- основные понятия и методы аналитической геометрии;
- основные понятия и методы линейной алгебры и теории алгебраических систем;
- основные понятия информатики,

уметь:

- использовать математические методы и модели для решения прикладных задач;
- использовать программные и аппаратные средства персонального компьютера.

владеть:

- владеть методами количественного анализа процессов обработки, поиска и передачи информации;
- навыками поиска информации в глобальной информационной сети Интернет и работы с офисными приложениями (текстовыми процессорами, электронными таблицами, средствами подготовки презентационных материалов, СУБД и т.п.).

2.3. Последующие учебные дисциплины (модули) и (или) практики, для которых необходимы знания, умения, навыки, формируемые данной учебной дисциплиной (модулем):

- Сети и системы передачи информации;
- Методы и средства криптографической защиты информации;
- Анализ и оценка рисков;
- Анализ данных в информационной безопасности.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

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 использовать необходимые математические методы для решения задач профессиональной деятельности.

Таблица 1 – Декомпозиция результатов обучения

Код и наименование компетенции	Планируемые результаты обучения по дисциплине (модулю)		
	Знать (1)	Уметь (2)	Владеть (3)
ОПК-3. Способен использовать необходимые математические методы для решения задач профессиональной деятельности	ИОПК-3.1. Знать: основы математики, основные математические методы.	ИОПК-3.2. Уметь: решать стандартные профессиональные задачи с применением методов математического анализа и моделирования.	ИОПК-3.3. Владеть: навыками математического исследования объектов профессиональной деятельности.

4. Структура и содержание дисциплины (модуля)

Объем дисциплины (модуля) в зачетных единицах **3 зачетные единицы**. Всего 108 часов: 36 часов выделено на контактную работу обучающихся с преподавателем (лекции – 18 часов, лабораторные работы – 18 часов), 72 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	П	ЛР	КР	СР	
1.	Раздел № 1. Основы Python и анализа данных.	3	1-4	4		4		14	Лабораторная работа 1
2.	Раздел № 2. Предобработка данных	3	5-7	3		3		13	Лабораторная работа 2. Контрольная работа 1
3.	Раздел № 3. Исследовательский анализ данных	3	8-11	4		4		14	Лабораторная работа 3
4.	Раздел № 4. Статистический анализ данных	3	12-14	3		3		13	Лабораторная работа 4
5.	Раздел № 5. Теория вероятностей	3	15-18	4		4		18	Лабораторная работа 5. Тест, вопросы к

№ 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) зачету
				Л	П	ЛР	КР	СР	
10	Итого	108		18		18		72	зачет

Примечание: Л – лекция; ПЗ – практическое занятие, семинар; ЛР – лабораторная работа; КР – курсовая работа; СР – самостоятельная работа.

Таблица 3 – Матрица соотнесения тем/разделов учебной дисциплины/модуля и формируемых компетенций

Темы, разделы дисциплины	Кол-во часов	Код компетенции (указываются компетенции перечисленные в п.3)	общее количество компетенций
		ОПК-3	
Раздел № 1. Основы Python и анализа данных.	22	+	1
Раздел № 2. Предобработка данных	19	+	1
Раздел № 3. Исследовательский анализ данных	22	+	1
Раздел № 4. Статистический анализ данных	19	+	1
Раздел № 5. Теория вероятностей	26	+	1
Итого	108		1

Содержание дисциплины

Раздел № 1. Основы Python и анализа данных.

Введение и основы синтаксиса. Списки и циклы. Операции с таблицами. Условия и функции. Библиотека Pandas. Предобработка данных. Анализ данных и оформление результатов. Веб-оболочка Jupyter Notebook. Процесс анализа данных

Раздел № 2. Предобработка данных

Основные принципы предобработки. Работа с пропусками. Изменение типов. Поиск дубликатов. Категоризация данных.

Раздел № 3. Исследовательский анализ данных

Введение в исследовательский анализ. Графики и выводы. Изучение срезов данных. Работа с несколькими источниками данных. Взаимосвязь данных. Валидация данных

Раздел № 4. Статистический анализ данных

Введение в статистический анализ. Описательная статистика. Математическая статистика. Выборочные методы математической статистики. Основы теории оценивания.

Основы проверки статистических гипотез. Основы регрессионного анализа.

Раздел № 5. Теория вероятностей

Случайные события. Пространство элементарных событий. Статистическая вероятность. Условные вероятности. Последовательности независимых испытаний. Случайные величины. Функции и числовые характеристики случайных величин. Основные законы распределения случайных величин. Экспоненциальное распределение и его числовые характеристики. Предельные теоремы теории вероятностей.

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРЕПОДАВАНИЮ И ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

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

При подготовке к лекционным и практическим занятиям необходимо воспользоваться учебно-методической литературой из п.8. Лекции необходимо проводить с использованием презентаций, созданных в прикладном пакете Microsoft Office PowerPoint.

При подготовке к лабораторным занятиям необходимо воспользоваться учебно-методической литературой (основной и дополнительной) и информационно-справочными ресурсами из п.8.

Тестирование допускается проводить в бумажном или электронном виде на специализированных образовательных площадках (Moodle).

5.2. Указания для обучающихся по освоению дисциплины (модулю)

Таблица 4 – Содержание самостоятельной работы обучающихся

<i>Номер раздела (темы)</i>	<i>Темы/вопросы, выносимые на самостоятельное изучение</i>	<i>Кол-во часов</i>	<i>Формы работы</i>
1	Раздел № 1. Основы Python и анализа данных.	14	Внеаудиторная, изучение учебных пособий
2	Раздел № 2. Предобработка данных	13	Внеаудиторная, изучение учебных пособий
3	Раздел № 3. Исследовательский анализ данных	14	Внеаудиторная, изучение учебных пособий
4	Раздел № 4. Статистический анализ данных	13	Внеаудиторная, изучение учебных пособий
5	Раздел № 5. Теория вероятностей	18	Внеаудиторная, изучение учебных пособий

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

В процессе изучения дисциплины предусмотрены следующие виды и формы

письменных работ для самостоятельного выполнения:

- 1) аудиторная контрольная работа;
- 2) типовой расчет - внеаудиторная работа;
- 3) домашнее задание, как теоретического, так и практического характера;
- 4) лабораторная работа, полностью или частично выполняемая в математических пакетах;
- 5) зачетная работа.

Контрольные работы и зачетная работа выполняется студентом в аудитории. Типовой расчет выполняется вне аудитории за определенный промежуток времени, установленный преподавателем, оформляется в отдельной тетради. В установленный срок студент сдает типовой расчет и устно отчитывается преподавателю по выполненной работе. Зачетную работу студент сдает в том случае, если по результатам семестрового контроля он набрал рейтинг менее 60 баллов. Лабораторные работы выполняются в аудитории и оцениваются как зачет/незачет. Для получения зачета по дисциплине по результатам текущего семестрового контроля студенту необходимо иметь рейтинг по контрольным работам не менее 60 баллов и зачет по всем лабораторным работам.

6. ОБРАЗОВАТЕЛЬНЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

6.1. Образовательные технологии

Раздел, тема дисциплины (модуля)	Форма учебного занятия		
	Лекция	Практическое занятие, семинар	Лабораторная работа
Раздел № 1. Основы Python и анализа данных.	Обзорная лекция	Не предусмотрено	выполнение контрольной работы, выполнение лабораторной работы
Раздел № 2. Предобработка данных	Лекция-диалог	Не предусмотрено	выполнение контрольной работы, выполнение лабораторной работы
Раздел № 3. Исследовательских анализ данных	Лекция	Не предусмотрено	выполнение контрольной работы, выполнение лабораторной работы,
Раздел № 4. Статистический анализ данных	Лекция	Не предусмотрено	выполнение контрольной работы, выполнение лабораторной работы
Раздел № 5. Теория вероятностей	Обзорная лекция	Не предусмотрено	выполнение контрольной

			работы, выполнение лабораторной работы
--	--	--	---

Учебные занятия по дисциплине могут проводиться с применением информационно-телекоммуникационных сетей при опосредованном (на расстоянии) интерактивном взаимодействии обучающихся и преподавателя в режимах on-line в формах: видеолекций, лекций-презентаций, видеоконференции, собеседования в режиме чат, форума, чата, выполнения виртуальных практических и/или лабораторных работ и др.

Максимальный объем занятий обучающегося с применением электронных образовательных технологий не должен превышать 25%.

6.2. Информационные технологии

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

- использование возможностей Интернета в учебном процессе (использование информационного сайта преподавателя (рассылка заданий, предоставление выполненных работ, ответы на вопросы, ознакомление учащихся с оценками и т.д.));
- использование электронных учебников и различных сайтов (например, электронные библиотеки, журналы и т.д.) как источников информации;
- использование возможностей электронной почты преподавателя;
- использование средств представления учебной информации (электронных учебных пособий и практикумов, применение новых технологий для проведения очных (традиционных) лекций и семинаров с использованием презентаций и т.д.);
- 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);
- использование виртуальной обучающей среды (или системы управления обучением LMS Moodle «Электронное образование») или иных информационных систем, сервисов и мессенджеров.

6.3. Программное обеспечение, современные профессиональные базы данных и информационные справочные системы

6.3.1. Программное обеспечение:

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

6.3.2. 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 ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

7.1. Паспорт фонда оценочных средств

При проведении текущего контроля и промежуточной аттестации по дисциплине (модулю) «Вероятностно-статистические методы в анализе 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 (модуля) – последовательным достижением результатов освоения содержательно связанных между собой разделов, тем.

Таблица 6 – Соответствие разделов, тем дисциплины (модуля), результатов обучения по дисциплине (модулю) и оценочных средств

№ п/п	Контролируемые разделы (темы) дисциплины*	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
1.	Раздел № 1. Основы Python и анализа данных.	ОПК-3	Лабораторная работа 1.
2.	Раздел № 2. Предобработка данных	ОПК-3	Лабораторная работа 2. Контрольная работа 1
3.	Раздел № 3. Исследовательских анализ данных	ОПК-3	Лабораторная работа 3.
4.	Раздел № 4. Статистический анализ данных	ОПК-3	Лабораторная работа 4.
5.	Раздел № 5. Теория вероятностей	ОПК-3	Лабораторная работа 5. Тест, вопросы к зачету

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

Для оценки результатов обучения применяются следующие критерии:

Таблица 7 – Показатели оценивания результатов обучения в виде знаний

Шкала оценивания	Критерии оценивания
5	демонстрирует глубокое знание теоретического материала, умение

«отлично»	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 8 – 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 и иные материалы, необходимые для оценки результатов обучения по дисциплине (модулю)

Раздел № 1. Основы Python и анализа данных.

Лабораторная работа 1. Библиотека NumPy

Цель занятия:

Получение представления о функциональности, доступных методах и объектах библиотеки NumPy. Изучение основных принципов практической работы с ними.

1. Создание массивов

В NumPy существует много способов создать массив. Один из наиболее простых – создать массив из обычных списков или кортежей Python, используя функцию `numpy.array()` (обратите внимание: `array()` – это метод, создающий объект типа `ndarray`):

```
In [1]: import numpy as np

a = np.array([1, 2, 3])
print(a)
print(type(a)) #вывод типа переменной a

[1 2 3]
<class 'numpy.ndarray'>
```

2. Базовые операции

Математические операции над массивами выполняются поэлементно. Ниже представлены примеры математических операций между двумя массивами.

```
In [19]: import numpy as np

a = np.array([20, 30, 40, 50])
b = np.arange(4)

print('a=',a)
print('b=',b)

print("n1", a + b)
print("2", a - b)
print("3", a * b)
print("4", a / b) # При делении на 0 возвращается inf (бесконечность)
print("5", a ** b) # Возведение в степень
print("6", a % b) # При взятии остатка от деления на 0 возвращается
```

3. Индексы, срезы, итерации

Одномерные массивы осуществляют операции индексирования, срезов и итераций очень схожим образом с обычными списками и другими последовательностями Python.

```
In [28]: a = np.arange(10) ** 3
print(np.arange(10))
print(a)

[0 1 2 3 4 5 6 7 8 9]
[ 0  1  8 27 64 125 216 343 512 729]
```

```
In [29]: a = np.arange(10) ** 3
print("1)", a)
print("2)", a[1])
print("3)", a[3:7])

a[3:7] = 8
print("4)", a)

print("5)", a[::-1])

for i in a:
    print(i ** (1/3))
```

4. Манипуляции с формой

У массива есть форма `shape`, определяемая числом элементов вдоль каждой оси:

```
In [49]: a = np.array([[0, 1, 2], [10, 12, 13]], [[100, 101, 102], [110, 112, 113]])
print ('a=',a)
print (a.shape) #вывод "формы", т.е. размерности

a= [[[ 0  1  2]
      [10 12 13]]

      [[100 101 102]
      [110 112 113]]]
(2, 2, 3)
```

5.Объединение массивов

Несколько массивов могут быть объединены вместе вдоль разных осей с помощью методов `hstack()` и `vstack()`. `hstack()` объединяет массивы по первым осям, `vstack()` – по последним.

6. Разбиение массива

Используя `hsplit()` вы можете разбить массив вдоль горизонтальной оси, указав либо число возвращаемых массивов одинаковой формы, либо номера столбцов, после которых массив разрежется «ножницами». Функция `vsplit()` разбивает массив вдоль вертикальной оси, а `array_split()` позволяет указать оси, вдоль которых произойдет разбиение.

Задания для лабораторной работы № 1:

1. Создание массива, заполненного нулями – приведите пример кода.
2. Математические операции между массивами. Представьте код, где массив со значениями от 1 до 9 умножается на константу 2. Должна получится таблица умножения для числа 2.
3. Слайсы (обрезка массива) – предоставите код для четырех случаев:
 - вывести все элементы, кроме первых трех;
 - вывести все элементы, кроме последних трех;
 - вывести все элементы, кроме первых и последних трех;
 - слайсы с двумерными массивами.

Раздел № 2. Предобработка данных

Лабораторная работа 2. Библиотека Pandas

Цель занятия:

Получение представления о функциональности, доступных методах и объектах библиотеки `pandas`. Изучение основных принципов практической работы с ними

1. Создание Series

Структура данных Series

Конструктор класса `Series` выглядит следующим образом:

```
pandas.Series(data=None, index=None, dtype=None, name=None, copy=False, fastpath=False)
```

```
In [2]: s1 = pd.Series([1, 2, 3, 4, 5])
print(s1)
```

```
0    1
1    2
2    3
3    4
4    5
dtype: int64
```

2. Создание DataFrame

Структура данных DataFrame

Конструктор класса `DataFrame` выглядит так:

```
pandas.DataFrame(data=None, index=None, columns=None, dtype=None, copy=False)
```

```
In [13]: nda1 = np.array([[1, 2, 3], [10, 20, 30]])
df4 = pd.DataFrame(nda1)
print(df4)
```

```
   0  1  2
0  1  2  3
1 10 20 30
```

3. Доступ к данным в структурах pandas

Основные методы для работы с объектами `DataFrame` представлены в таблице ниже:

Операция	Синтаксис	Возвращаемый результат
Выбор столбца	<code>df[col]</code>	<code>Series</code>
Выбор строки по метке	<code>df.loc[label]</code>	<code>Series</code>
Выбор строки по индексу	<code>df.iloc[loc]</code>	<code>Series</code>
Слайс по строкам	<code>df[0:4]</code>	<code>DataFrame</code>
Выбор строк, отвечающих условию	<code>df[bool_vec]</code>	<code>DataFrame</code>

Для доступа к данным можно использовать атрибуты структур, в качестве которых выступают метки.

4. Добавление элементов в структуры

Увеличение размера структуры – т.е. добавление новых, дополнительных, элементов – это довольно распространенная задача. Добавление нового элемента в `Series` показано ниже.

```
In [21]: s = pd.Series([10, 20, 30, 40, 50], ['a', 'b', 'c', 'd', 'e'])
print ('0) Исходный Series s\n', s)
s['f'] = 60
print ('1) Добавили s['f'] = 60\n', s)
```

5. Pandas и отсутствующие данные

Метод `isnull()` из библиотеки `pandas`, который позволяет быстро определить наличие элементов `NaN` в структурах.

Подробную информацию об объекте (датафрейме), для этого можно воспользоваться методом `info()`.

6. Замена отсутствующих данных

Отсутствующие данные объектов можно заменить на конкретные числовые значения, для этого можно использовать метод `fillna()`.

Не всегда целесообразно заменять отсутствующие данные нулями. Иногда недостающие данные можно заполнить средним значением по столбцу:

7. Удаление объектов/столбцов с отсутствующими данными

Довольно часто используемый подход при работе с отсутствующими данными – это удаление записей (строк) или полей (столбцов), в которых встречаются пропуски. Для того, чтобы удалить все объекты, которые содержат значения `NaN` воспользуйтесь методом `dropna()` без аргументов.

Задания для лабораторной работы № 2:

1. Приведите пример `Series` и попробуйте: выводить на экран, добавлять элементы, изменять значения.

2. Повторите предыдущий пункт для объекта `DataFrame`: выводить на экран, добавлять элементы, изменять значения.

3. Открыть при помощи `pandas` созданный в процессе выполнения работы `dataframe.csv` файл с пропусками или некорректными значениями и исправить их.

Контрольная работа 1

1. Перечислите наиболее важные атрибуты объектов `ndarray`
2. Базовая функциональность `NumPy` – опишите основные доступные функции.
3. Настройка печати массивов «под свои нужды».
4. Опишите применение унарных операций к массивам, приведите примеры.
5. Копии и представления при работе с массивами.
6. Сохранение массива в файл и чтение из файла.
7. Сравнение библиотеки `Pandas` с функциональностью библиотеки `NumPy`.

8. Опишите Конструктор класса `Series`.
9. Опишите Конструктор класса `DataFrame`.
10. Опишите сохранение и чтение объектов из файла.

Раздел № 3. Исследовательских анализ данных

Лабораторная работа 3. Статистика в NumPy и в Statistics. Введение в Scipy.Stats

Цель занятия:

Получение представления о функциональности, доступных методах и объектов статистики библиотек/модулей `NumPy`, `Statistics` и `Scipy.Stats`. Изучение основных принципов практической работы с ними.

1. Статистика в NumPy

Максимум и минимум массива

Определение минимального и максимального значения элементов массива вдоль указанной оси при помощи методов `np.amin()` и `np.amax()`.

Среднее арифметическое

Среднее арифметическое (или обычно говорят кратко – «среднее») – это сумма элементов, деленная на их сумму. Метод `np.mean()` вычисляет среднее значение, складывая все элементы массивов, а затем делит его на количество элементов. Мы также можем указать ось, по которой может быть вычислено среднее значение.

Медиана

Медиана (с лат. «середина» – середина множества чисел. Формула отличается для множеств с четным и нечетным числом элементов. Метод `np.median()` все отсортирует сам и найдет медиану.

Среднеквадратичное (стандартное) отклонение

Среднеквадратичное отклонение (формула 4, так же называется стандартным отклонением) – определяется как квадратный корень из дисперсии случайной величины. При помощи метода `np.std()` вычисляется стандартное отклонение («std» – от англ. standart, стандартное).

Среднее взвешенное

Функция `np.average()` определяет средневзвешенное значение, и может использоваться для многомерных массивов. Средневзвешенное значение рассчитывается путем умножения компонента на его вес, веса указываются отдельно. Если веса не указаны, результат будет таким же, как и среднее значение (среднее арифметическое).

Процентили

Процентиль (персентиль или персентиль) – это такое значение X_p из массива (или в терминологии статистики это называется совокупностью значений), что значения p -й части совокупности меньше или равны этому значению X_p .

Полный размах

Функция `np.ptp()` полезна для определения размаха значений вдоль оси.

2. Некоторые статистические функции из statistics

Гармоническое среднее

Гармоническое среднее является обратным средним значением обратных величин всех элементов в наборе данных

Среднее геометрическое

Геометрическое среднее является n -й корнем произведения всех элементов x_i в наборе данных X .

3. Статистика при помощи модуля Scipy.Stats

Этот модуль содержит большое количество распределений вероятностей, сводную и частотную статистику, корреляционные функции и статистические тесты, маскированную статистику, ядерную оценку плотности, функциональность квази-Монте-Карло и многое другое.

Все статистические функции и доступные случайные величины находятся в суб-пакете `scipy.stats`.

Кумулятивная функция распределения (CDF).

Рассмотрим кумулятивную функцию распределения (CDF, с англ. Cumulative Distribution Function). Она показывает вероятность того, что целевое значение меньше указанного значения либо равно ему.

Функция получения значения процентиля (PPF)

PPF (расшифровывается с англ. как percent point function) – функция, используемая для получения значения, соответствующему определенному процентилю.

Задания для лабораторной работы № 3:

1. Примеры применения среднеарифметического, медианы, среднегеометрического и гармонического среднего, процентилей.
2. Представьте код, который строит графики кумулятивной функции распределения (cdf), функции получения процентилей (ppf) и функции плотности вероятности (pdf) для сгенерированного при помощи NumPy набора дискретных значений от -3 до 3 с дискретностью 0.1.
3. Проведите эксперименты с библиотекой для построения графиков `matplotlib`. Попробуйте добавить `label` для графиков, `xlabel/ylabel` для осей, `title` и для всего графика целиком. На следующем занятии будет представлено решение этих задач.

Раздел № 4. Статистический анализ данных

Лабораторная работа № 4

Цель занятия:

Изучение функциональности, доступных методов и объектов статистики библиотеки `Scipy` – генерация случайных чисел и параметры формы распределения. Исследование принадлежности ряда значений к конкретному закону распределения (на основе параметров формы, куртозиса и др.). Построение диаграмм размах. Рассмотрение примеров описательной статистики из реального мира (z -статистика и p -значение).

Генерация случайных чисел

Генерация случайных чисел (рядов величин случайного распределения) зависит от генераторов из модуля `numpy.random`.

Генерация по нормальному закону распределения. Генерация 5 случайных величин, распределенных по нормальному закону распределения, используя вызов `<norm.rvs(size=5, random_state=rng)>`.

Кроме генерации рядов с нормальным распределением, можно генерировать наборы

случайных величин, с гамма распределением, или с любым другим типом распределения. Полный список которых можно найти в официальной документации `scipy` или `numpy`.

Принцип такой же (как и при рассмотренном ранее вызове `norm.rvs()`), только может отличаться набор передаваемых параметров.

Куртозис и сдвиг (моменты, и прочие понятия статистики)

В `scipy` доступны функции и методы для получения и многих других, кроме уже рассмотренных, параметров статистики. Представлено вычисление куртозиса (острота пика в середине) и перекоса (значений относительно медианы).

Диаграмма размаха

Диаграммы размаха в удобной форме показывают медиану (или, если нужно, среднее), нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы. Расстояния между различными частями ящика так же позволяют определить степень разброса (дисперсии) и асимметрии данных.

Примеры статистики из жизни

Пример с временем доставки. Рассмотрим реальный пример из жизни: статистика времени доставки пиццы.

В течение года мы заказываем пиццу на дом, при этом мы каждый раз смотрим на настенные стрелочные часы, отмечая время, которое проходит между заказом и доставкой, целым количеством минут. Тогда накопленные за год данные о доставке пиццы могли бы выглядеть так:

```
norm_rv = stats.norm(
    loc=30, scale=5
) # генерируем норм расп значения возле 30 плюс минус 5
samples = np.trunc(norm_rv.rvs(365))
```

Теперь выясним среднее время доставки пиццы и его среднеквадратическое отклонение:

```
samples.mean(), samples.std()
```

Можно сказать, что время доставки пиццы занимает около 30 ± 5 минут. Посмотрим на то, как распределены данные:

```
sns.histplot(x=samples, discrete=True);
```

Вычислить на какую величину время доставки отличается от среднего времени доставки, Интерпретировать данный результат, используя Z -значение. Можно воспользоваться графиком:

```
fig, ax = plt.subplots()
x = np.linspace(norm_rv.ppf(0.001), norm_rv.ppf(0.999), 200)
ax.vlines(40.5, 0, 0.1, color="k", lw=2)
sns.lineplot(x=x, y=norm_rv.pdf(x), color="r", lw=3)
sns.histplot(x=samples, stat="probability", discrete=True);
```

Задания для лабораторной работы № 4:

1. Определить к какому конкретно типу распределения относится время доставки пиццы из множества всевозможных?

2. Постройте эмпирические функции распределения (относительные и накопленные частоты) для роста (в см) группы из 20 мужчин: 181, 169, 178, 178, 171, 179, 172, 181, 179,

168, 174, 167, 169, 171, 181, 181, 183, 172, 176, 181.

3. Построить эмпирические функции распределения (относительные и накопленные частоты) успеваемости в группе из 20 студентов: 4, 4, 5, 3, 4, 5, 4, 5, 3, 5, 3, 3, 5, 4, 5, 4, 3, 5, 3, 5.

Раздел № 5. Теория вероятностей

Лабораторная работа 5

Цель занятия:

Получение представления о функциональности, доступных методах и объектах библиотеки `scikit-learn`. Изучение основных принципов практической работы с ними для реализации математических алгоритмов. Рассмотрение основных положений машинного обучения с учителем и без учителя: метод k -средних, линейная регрессия и деревья решений.

Обучение с учителем

Обучение с учителем (англ. Supervised learning) – это один из способов машинного обучения, в ходе которого испытуемая система принудительно обучается с помощью примеров «стимул–реакция». К обучению с учителем относятся задачи классификации и регрессии.

Классификация – система группировки объектов исследования или наблюдения в соответствии с их общими признаками. При классификации происходит предсказание признака, множество допустимых значений которого ограничено.

Регрессия – выявление зависимости между случайными переменными и математическим выражением, отражающим связь между зависимой переменной y и независимыми переменными x_{i*} при условии, что это выражение будет иметь статистическую значимость. В отличие от чисто функциональной зависимости $y = f(x_{i*})$, когда каждому значению независимой переменной x_{i*} соответствует одно определённое значение величины y , при регрессионной связи одному и тому же значению x_{i*} могут соответствовать в зависимости от случая различные значения величины y .

Обучение без учителя

Обучение без учителя (самообучение, спонтанное обучение, англ. Unsupervised learning) — один из способов машинного обучения, при котором испытуемая система спонтанно обучается выполнять поставленную задачу без вмешательства со стороны экспериментатора.

Как правило, оно пригодно только для задач, в которых известны описания множества объектов (т.е. существуют обучающие выборки), и требуется обнаружить внутренние взаимосвязи или закономерности, существующие между объектами.

Примерами обучения без учителя являются:

- кластеризация – задача группировки множества объектов на подмножества (кластеры) таким образом, чтобы объекты из одного кластера были более похожи друг на друга, чем на объекты из других кластеров по какому-либо критерию.
- снижение размерности – представление данных в пространстве меньшей размерности с минимальными потерями полезной информации. Обычно в его основе лежит метод главных компонент).

- выявление аномалий – это опознавание во время интеллектуального анализа данных редких данных, событий или наблюдений, которые вызывают подозрения ввиду существенного отличия от большей части данных.

Функциональность Scikit-learn

Библиотека `scikit-learn` реализует следующие основные методы:

- **Линейные:** модели, задача которых построить разделяющую или аппроксимирующую гиперплоскость (для классификации и регрессии соответственно).
- **Метрические:** модели, которые вычисляют расстояние по одной из метрик между объектами выборки, и принимают решения в зависимости от этого расстояния (например, метод К-ближайших соседей).
- **Деревья решений:** обучение моделей, базирующихся на множестве условий, оптимально выбранных для решения задачи.
- **Ансамблевые методы:** методы, основанные на деревьях решений, которые комбинируют мощь множества деревьев, и таким образом повышают их качество работы, а также позволяют производить отбор признаков (бустинг, бэггинг, случайный лес, мажоритарное голосование).
- **Нейронные сети:** комплексный нелинейный метод для задач регрессии и классификации.
- **Метод опорных векторов** (англ. support vector machine, SVM): нелинейный метод, который обучается определять границы принятия решений.
- **Наивный Байес:** прямое вероятностное моделирование для задач классификации.
- **Метод главных компонент** (англ. principal component analysis, PCA): линейный метод понижения размерности и отбора признаков.
- **Стохастическое вложение соседей с t-распределением** (англ. t-distributed Stochastic Neighbor Embedding, t-SNE): нелинейный метод понижения размерности.
- **К-средних:** самый распространенный метод для кластеризации, требующий на вход число кластеров, по которым должны быть распределены данные.
- **Кросс-валидация:** метод, при котором для обучения используется весь датасет (в отличие от разбиения на выборки `train/test`), однако обучение происходит многократно, и в качестве валидационной выборки на каждом шаге выступают разные части датасета. Итоговый результат является усреднением полученных результатов.
- **Поиск по сетке** (англ. Grid Search): метод для нахождения оптимальных гиперпараметров[2] модели путем построения сетки из значений гиперпараметров и последовательного обучения моделей со всеми возможными комбинациями гиперпараметров из сетки.

Задания для лабораторной работы № 5:

1. Кластерный анализ – что такое и где применяется, чем отличается от классификации.
2. Линейная регрессия – проведите эксперимент с собственным датасетом (к примеру, подготовьте в `Excel` датасет случайных значений, лежащих возле какого-либо уравнения).

Перечень вопросов к зачету

1. Основы Python и анализа данных.
2. Введение и основы синтаксиса.
3. Списки и циклы.
4. Операции с таблицами.
5. Условия и функции.
6. Библиотека Numpy.
7. Библиотека Pandas.
8. Веб-оболочка Jupyter Notebook.
9. Анализ данных и оформление результатов.
10. Процесс анализа данных.
11. Предобработка данных. Основные принципы предобработки.
12. Предобработка данных. Работа с пропусками.
13. Предобработка данных. Изменение типов.
14. Предобработка данных. Поиск дубликатов.
15. Предобработка данных. Категоризация данных.
16. Исследовательский анализ данных. Графики и выводы.
17. Исследовательский анализ данных. Изучение срезов данных.
18. Исследовательский анализ данных. Работа с несколькими источниками данных.
19. Исследовательский анализ данных. Взаимосвязь данных.
20. Исследовательский анализ данных. Валидация данных.
21. Статистический анализ данных. Описательная статистика.
22. Математическая статистика.
23. Выборочные методы математической статистики.
24. Основы теории оценивания.
25. Основы проверки статистических гипотез.
26. Основы регрессионного анализа.
27. Теория вероятностей.
28. Случайные события.
29. Пространство элементарных событий.
30. Статистическая вероятность.
31. Условные вероятности.
32. Последовательности независимых испытаний.
33. Случайные величины.
34. Функции и числовые характеристики случайных величин.
35. Основные законы распределения случайных величин.
36. Экспоненциальное распределение и его числовые характеристики.
37. Предельные теоремы теории вероятностей.

Таблица 9 – Примеры оценочных средств с ключами правильных ответов

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
ОПК-3. Способен использовать необходимые математические методы для решения задач профессиональной деятельности				
1.	Задание закрытого типа	На плоскости нарисованы две концентрические окружности, радиусы которых 6 и 12 см соответственно. Какова вероятность того, что точка брошенная наудачу в большой круг, попадет в кольцо, образованное указанными окружностями? 1) 0,5; 2) 0,65; 3) 0,12; 4) 0,75; 5) 0,60.	4	2
2.		На сборку попадают детали с двух автоматов: 80 % из первого и 20 % из второго. Первый автомат дает 10 % брака, второй – 5 % брака. Найти вероятность попадания на сборку доброкачественной детали. 1) 0,90; 2) 0,09; 3) 0,91; 4) 0,85; 5) 0,15	3	3
3.		Случайная величина распределена по нормальному закону, причем $M(X) = 15$. Найти $P(10 < X < 15)$, если известно, что $P(15 < X < 20) = 0,25$. 1) 0,10; 2) 0,15; 3) 0,20; 4) 0,25; 5) 0,30.	4	3
4.		По выборке объема $n = 51$ найдена смещенная оценка $Dv=3$ генеральной дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности. 1) 3,05; 2) 3,06; 3) 3,51; 4) 3,60; 5) 0.	2	3
5.		Как называется число m_0 (наступления события в n независимых испытаниях, в каждом из которых вероятность появления события равна p), определяемое из неравенства: $np - q \leq m_0 \leq np + p$ 1) наибольшее 2) оптимальное 3) наивероятнейшее 4) невозможное 5) минимальное	3	3

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)										
6.	Задание открытого типа	Задан статистический ряд распределения <table border="1" style="margin-left: 20px;"> <tr> <td>Варианта x_i</td> <td>1</td> <td>2</td> <td>5</td> <td>7</td> </tr> <tr> <td>Частота n_i</td> <td>10</td> <td>50</td> <td>25</td> <td>15</td> </tr> </table> <p>Найти выборочную среднюю \bar{X}. В ответ записать число $5\bar{X}$.</p>	Варианта x_i	1	2	5	7	Частота n_i	10	50	25	15	17	8
Варианта x_i		1	2	5	7									
Частота n_i		10	50	25	15									
7.		Данное предприятие в среднем выпускает 20 % продукции высшего сорта и 70 % продукции первого сорта. Найти вероятность P того, что случайно взятое изделие этого предприятия будет высшего или первого сорта. В ответ записать число $30P$.	27	6										
8.		Случайная величина X распределена равномерно на интервале $(2; 6)$ и $p(x)$ – ее плотность вероятности. Найти $p(3)$. В ответ записать число $40p(3)$.	10	8										
9.	Студентам нужно сдать 4 экзамена за 6 дней. Сколькими способами можно составить расписание сдачи экзаменов?	360	8											
10.	Вероятность появления события A в каждом из 100 независимых испытаний равна 0,4. Найдите математическое ожидание и дисперсию случайной величины X – числа появлений события A . В ответ запишите их сумму.	64	8											

Полный комплект оценочных материалов по дисциплине (модулю) (фонд оценочных средств) хранится в электронном виде на кафедре, утверждающей рабочую программу дисциплины (модуля), и в Центре мониторинга и аудита качества обучения.

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

По дисциплине «Вероятностно-статистические методы в анализе данных» итоговой формой отчетности в четвертом семестре является зачет. Согласно действующей в АГУ системе оценивания БАРС для дисциплин, итоговой формой отчетности для которых является зачет, отводится 100 баллов (90 баллов на текущие формы контроля и до 10 баллов отводится на бонусы), которые накапливаются студентом в течение всего семестра изучения дисциплины и распределяются по возможности равномерно по всему семестру. Система накопления баллов, а также система штрафов, представлена в технологической карте дисциплины «Вероятностно-статистические методы в анализе данных».

Отчет по лабораторной работе

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:

- отсутствия необходимых разделов,
- отсутствия необходимого графического материала,
- неверных результатов расчета.

Контрольные работы

Контрольная работа состоит из 2-х заданий.

Основаниями для снижения оценки за задание являются:

- ошибки в объяснениях и комментариях при верно выполненном задании;
- неполный ответ для теоретических заданий;
- небрежное выполнение;
- многократное переписывание контрольной работы.

Задание не может быть засчитано, если:

- даны два неверных ответа на теоретические вопросы.

В соответствии с балльно-рейтинговой системой БАРС по дисциплине отводится 100 баллов (90 баллов на текущие формы контроля и до 10 баллов отводится на бонусы), которые накапливаются студентом в течение всего семестра изучения дисциплины.

Оценивание студентов на зачете осуществляется в соответствии с требованиями и критериями 100-балльной шкалы. Учитываются как результаты текущего контроля, так и знания, навыки и умения, непосредственно показанные студентами в ходе зачета.

Текущий контроль осуществляется в ходе учебного процесса и консультирования студентов, по результатам выполнения самостоятельных и тематических контрольных 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 зачете осуществляется комплексная проверка знаний, навыков и умений студентов по всему теоретическому материалу дисциплины и с проверкой практических навыков и умений по разработке документов различных видов. Теоретические знания оцениваются путем компьютерного тестирования или на основании письменных ответов студентов по нескольким теоретическим вопросам.

Таблица 10 – Технологическая карта рейтинговых баллов по дисциплине (модулю)

№ п/п	Контролируемые мероприятия	Количество мероприятий / баллы	Максимальное количество баллов	Срок представления
Основной блок				
1.	<i>Выполнение лабораторной работы</i>	5/15	75	По расписанию
2.	<i>Выполнение контрольной работы</i>	1/10	10	
3.	<i>Тест</i>	1/5	5	
Всего			90	-
Блок бонусов				
4.	<i>Посещение занятий без пропусков</i>	1	3	
5.	<i>Своевременное выполнение всех заданий</i>	1	3	
6.	<i>Активность студента на занятии</i>	1	4	
Всего			10	-
ИТОГО			100	-

Таблица 11 – Система штрафов (для одного занятия)

Показатель	Балл
<i>Опоздание на занятие</i>	- 1
<i>Нарушение учебной дисциплины</i>	- 1
<i>Неготовность к занятию</i>	- 2
<i>Пропуск занятия без уважительной причины</i>	- 2

Таблица 12 – Шкала перевода рейтинговых баллов в итоговую оценку за семестр по дисциплине (модулю)

Сумма баллов	Оценка по 4-балльной шкале	
90–100	5 (отлично)	Зачтено
85–89	4 (хорошо)	
75–84		
70–74		
65–69	3 (удовлетворительно)	
60–64		
Ниже 60	2 (неудовлетворительно)	Не зачтено

При реализации дисциплины (модуля) в зависимости от уровня подготовленности обучающихся могут быть использованы иные формы, методы контроля и оценочные средства, исходя из конкретной ситуации.

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

8.1. Основная литература:

1. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2020. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450262>
2. Теория вероятностей и математическая статистика. Практикум [Электронный ресурс]: учеб. пособие / И.Ю. Мацкевич, Н.П. Петрова, Л.И. Тарусина - Минск : РИПО, 2017. - URL: <http://www.studentlibrary.ru/book/ISBN9789855037119.html> (ЭБС «Консультант студента»).
3. Ракитский, А. А. Методы машинного обучения : учебно-методическое пособие / А. А. Ракитский. — Новосибирск : Сибирский государственный университет телекоммуникаций и информатики, 2018. — 32 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/90591> .

4. Рашка С., Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения / Рашка С. - М. : ДМК Пресс, 2017. - 418 с. - ISBN 978-5-97060-409-0 - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970604090.html>

5. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2020. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450166>

8.2. Дополнительная литература:

1. Теория вероятностей. Примеры и задачи [Электронный ресурс]: учеб. пособие / Васильчик М.Ю. - Новосибирск : Изд-во НГТУ, 2014. - URL: <http://www.studentlibrary.ru/book/ISBN9785778224872.html> (ЭБС «Консультант студента»).

2. Теория вероятностей и математическая статистика [Электронный ресурс] / Балдин К. В. - М. : Дашков и К, 2014. - URL: <http://www.studentlibrary.ru/book/ISBN9785394021084.html> (ЭБС «Консультант студента»).

3. Федин, Ф. О. Анализ данных. Часть 1. Подготовка данных к анализу : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 204 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/26444.html>

4. Билл, Фрэнкс Укрощение больших данных : как извлекать знания из массивов информации с помощью глубокой аналитики / Фрэнкс Билл ; перевод А. Баранов. — Москва : Манн, Иванов и Фербер, 2014. — 340 с. — ISBN 978-5-00057-146-0. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/39433.html>

5. Флах П., Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / Флах П. - М. : ДМК Пресс, 2015. - 400 с. - ISBN 978-5-97060-273-7 - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970602737.html>

8.3. Интернет-ресурсы, необходимые для освоения дисциплины (модуля)

1 Электронно-библиотечная система (ЭБС) ООО «Политехресурс» «Консультант студента». www.studentlibrary.ru.

2 Электронно-библиотечная система IPR BOOKS <http://www.iprbookshop.ru/>

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Для проведения лабораторных занятий необходима компьютерная аудитория, в которой организован доступ к сети Интернет и установлено программное обеспечение. Для проведения публичной защиты проектов, необходима мультимедийная аудитория с проектором.

Во время лекций используется ноутбук и проектор для презентаций. Семинары проходят в компьютерном классе. Слушателям курса рекомендуется использовать свои ноутбуки с установленной последней версией Anaconda для Python 3.7.

Рабочая программа дисциплины (модуля) при необходимости может быть адаптирована для обучения (в том числе с применением дистанционных образовательных технологий) лиц с ограниченными возможностями здоровья, инвалидов. Для этого требуется заявление обучающихся, являющихся лицами с ограниченными возможностями здоровья, инвалидами, или их законных представителей и рекомендации психолого-

медико-педагогической комиссии. Для инвалидов содержание рабочей программы дисциплины (модуля) может определяться также в соответствии с индивидуальной программой реабилитации инвалида (при наличии).