

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  
«05» мая 2025 г.

УТВЕРЖДАЮ  
И.о. заведующего кафедрой  
информационной безопасности

В.А. Черкасова  
«05» мая 2025 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**  
**АНАЛИЗ ДАННЫХ В ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ**  
*наименование*

Составитель(-и)

**Мартьянова А.Е., к.т.н., доцент кафедры  
информационной безопасности  
10.03.01 Информационная безопасность**

Направление подготовки /  
специальность

Направленность (профиль) ОПОП

**«Организация и технология защиты информации  
(в сфере информационных и  
телекоммуникационных технологий)»**

Квалификация (степень)

**бакалавр**

Форма обучения

**очно-заочная**

Год приема

**2023**

Курс

**5**

Семестр

**9**

## 1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

**1.1. Целями освоения дисциплины (модуля)** «Анализ данных в информационной безопасности» является развитие навыков программирования на языке Python, знакомство с понятиями о сборе, обработке и анализе данных в сфере информационной безопасности, введение в автоматизированные методы работы с данными - машинное обучение и нейронные сети.

### 1.2. Задачи освоения дисциплины (модуля):

- освоение продвинутых методов исследования взаимосвязей между показателями, характеризующими объекты в социально-экономических исследованиях;
- освоение продвинутых методов распознавания образов и типологизации объектов;
- освоение продвинутых методов оптимизации представления информации об объектах;
- освоение современных пакетов прикладных программ, реализующих алгоритмы многомерного анализа данных;
- приобретение навыков содержательной интерпретации результатов исследования.

## 2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОПОП

**2.1. Учебная дисциплина (модуль)** Б1.В.Д.05.02 «Анализ данных в информационной безопасности» относится к блоку элективных дисциплин учебного плана направления подготовки 10.03.01 ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ, профиль «Организация и технология защиты информации» 2023 года набора и осваивается в 9 семестре, общая трудоемкость дисциплины — 3 ЗЕ, 108 часов, итоговая форма контроля — экзамен.

Дисциплина «Анализ данных в информационной безопасности» встраивается в структуру ОПОП (последовательность в учебном плане) как с точки зрения преемственности содержания, так и с точки зрения непрерывности процесса формирования компетенций выпускника. Обязательными знаниями для успешного освоения курса является основы математической статистики и теории вероятности, основы программирования на языке Python.

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

- Вероятностно-статистические методы в анализе данных.
- Основы программирования.

#### знания:

- основ математики, основных математических методов,
- основ программирования;

#### умения:

- решать стандартные профессиональные задачи с применением методов математического анализа и моделирования,
- использовать языки программирования и технологии программных средств для решения задач профессиональной деятельности;

#### навыки:

- навыками математического исследования объектов профессиональной деятельности,
- навыками программирования для решения задач профессиональной деятельности.

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

- Бакалаврская работа.

### 3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

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

а) профессиональных (ПК):

- способен выполнять работы по установке, настройке и техническому обслуживанию защищенных технических средств обработки информации (ПК-2);

- способен осуществлять внедрение систем защиты информации для обеспечения информационной безопасности автоматизированных систем (ПК-3).

**Таблица 1 – Декомпозиция результатов обучения**

Код компетенции и наименование	Планируемые результаты обучения по дисциплине (модулю)		
	Знать	Уметь	Владеть
ПК-2. Способен выполнять работы по установке, настройке и техническому обслуживанию защищенных технических средств обработки информации	ИПК-2.1. Знать: технические описания и инструкции по эксплуатации технических средств обработки информации в защищенном исполнении, методы контроля защищенности информации от несанкционированного доступа и специальных программных воздействий, порядок аттестации объектов информатизации на соответствие требованиям безопасности информации	ИПК-2.2. Уметь: проводить настройку защищенных технических средств обработки информации в соответствии с инструкциями по эксплуатации и эксплуатационно-техническими документами, проводить техническое обслуживание защищенных технических средств обработки информации в соответствии с инструкциями по эксплуатации и эксплуатационно-технической документацией.	ИПК-2.3. Владеть: методами защиты информации от несанкционированного доступа и специальных программных воздействий на нее
ПК-3. Способен осуществлять внедрение систем защиты информации для обеспечения информационной безопасности автоматизированных систем	ИПК-3.1. Знать: основные угрозы безопасности информации и модели нарушителя в автоматизированных системах, содержание эксплуатационной документации	ИПК-3.2. Уметь: администрировать программные средства системы защиты информации автоматизированных систем, устранять известные уязвимости автоматизированной системы,	ИПК-3.3. Владеть: методикой анализа структурных и функциональных схем защищенной автоматизированной системы

	автоматизированной системы, типовые средства, методы и протоколы идентификации, аутентификации и авторизации основные меры по защите информации в автоматизированных системах, нормативные правовые акты в области защиты информации	приводящие к возникновению угроз безопасности информации, применять аналитические и компьютерные модели автоматизированных систем и систем защиты информации, определять параметры настройки программного обеспечения системы защиты информации автоматизированной системы	
--	--	--	--

#### 4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Объем дисциплины (модуля) в зачетных единицах (3 зачетные единицы) 108 часов, на контактную работу обучающихся с преподавателем выделено 27 часов (из них лекции – 9 часов, лабораторных работ – 18 часов) и на самостоятельную работу обучающихся 81 час.

**Таблица 2 – Структура и содержание дисциплины (модуля)**

№ п/п	Наименование раздела, темы	Семестр	Неделя семестра	Контактная работа (в часах)			Самостоят. работа		Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
				Л	ПЗ	ЛР	КР	СР	
1	Основы программирования на языке Python.	9	1-2	1		2		11	Лабораторная работа №1
2	Математический аппарат.	9	3-4	1		2		11	Лабораторная работа №2
3	Введение в модуль NumPy.	9	5-6	1		2		11	Лабораторная работа №3
4	Основы работы с Pandas.	9	7-8	1		2		11	Лабораторная работа №4
5	Разведывательный анализ данных.	9	9-10	1		2		11	Лабораторная работа №5
6	Визуализация данных. Представление	9	11-14	2		4		12	Лабораторная работа №6. Практическая работа № 1

	результатов								
7	Работа с текстовыми данными. Текстовый анализ.	9	15-18	2		4		14	Лабораторная работа №7. Проверочный тест
<b>ИТОГО</b>			<b>108</b>	<b>9</b>		<b>18</b>		<b>81</b>	<b>ЭКЗАМЕН</b>

*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**

Разделы, темы дисциплины (модуля)	Кол-во часов	Компетенции		общее количество компетенций
		ПК-2	ПК-3	
Основы программирования на языке Python.	14	+	+	2
Математический аппарат.	14	+	+	2
Введение в модуль NumPy.	14	+	+	2
Основы работы с Pandas.	14	+	+	2
Разведывательный анализ данных.	14	+	+	2
Визуализация данных. Представление результатов	18	+	+	2
Работа с текстовыми данными. Текстовый анализ.	20	+	+	2
<b>Итого</b>	108			2

## СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

### Раздел 1. Основы программирования на языке Python

Введение в анализ данных на языке Python. Почему Python становится стандартом для работы с большими данными. Прикладные задачи политологов, для решения которых подходит язык Python. Основы программирования на языке Python (refresher): типы данных и методы работы с ними (переменные, листы, словари, кортежи).

### Раздел 2. Математический аппарат

Математический аппарат для анализа данных: векторы, матрицы, функции и производные. Основы программирования на языке Python(refresher): циклы функции, знание синтаксиса языка.

### Раздел 3. Введение в модуль NumPy

Введение в модуль для работы с числовыми данными NumPy (Numerical Python). Особенности типы данных в NumPy. Работа с векторами и матрицами. Вычисление главных статистических метрик с помощью NumPy (среднее, медиана, мода, дисперсия).

### Раздел 4. Основы работы с Pandas

Введение в модуль для работы с табличным представлением данных Pandas. Преобразование словарей в табличный формат Pandas, загрузка данных из внешних источников. Особенности фильтрации и обращения к данным. Работа с табличными данными в Pandas на примере.

### Раздел 5. Основные этапы анализа данных

Предобработка данных. Разведывательный анализ данных. Статистический анализ данных. Регрессионные модели. Метод наименьших квадратов. Логистическая регрессия. Проведение разведывательного анализа данных на примере.

## **Раздел 6. Визуализация данных. Представление результатов**

Введение в визуализацию данных. Нюансы визуализации данных и принципы человеческого восприятия. Правила создания хороших визуализаций. Создание различных видов визуализаций на синтетических данных и тренировочных наборах данных. Создание интерактивных визуализаций и отчетов с помощью библиотеки Plotly.

## **Раздел 7. Работа с текстовыми данными. Текстовый анализ**

Введение в анализ текста. Особенности подготовки данных для анализа текста. Применение в информационной безопасности.

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

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

Учебная деятельность студента в процессе изучения строится из контактных форм работы с преподавателем (аудиторные занятия, зачет, экзамен) и самостоятельной работы.

Для успешного освоения дисциплины является обязательным посещение всех занятий, выполнение домашнего задания и иных форм самостоятельной работы, которые назначаются преподавателем.

Аудиторные занятия проводятся на основе теоретического материала, опубликованного на образовательном портале, это позволяет студентам изучить пропущенный материал или самостоятельно разобраться с темой, не освоенной на занятии.

Для исключения отрыва студентов от учебного процесса проводится учет посещаемости аудиторных занятий.

### **5.2. Указания для обучающихся по освоению дисциплины (модулю)**

**Таблица 4 – Содержание самостоятельной работы обучающихся**

Номер раздела (темы)	Темы/вопросы, выносимые на самостоятельное изучение	Кол-во часов	Формы работы
Основы программирования на языке Python.	Подготовка отчета по лабораторной работе 1	11	Внеаудиторная, изучение учебных пособий и нормативных документов
Математический аппарат.	Подготовка отчета по лабораторной работе 2	11	Внеаудиторная, изучение учебных пособий и нормативных документов
Введение в модуль NumPy.	Подготовка отчета по лабораторной работе 3	11	Внеаудиторная, изучение учебных пособий и нормативных документов
Основы работы с Pandas.	Подготовка отчета по лабораторной работе 4	11	Внеаудиторная, изучение учебных пособий и нормативных документов
Разведывательный анализ данных.	Подготовка отчета по лабораторной работе 5	11	Внеаудиторная, изучение учебных пособий и

Номер раздела (темы)	Темы/вопросы, выносимые на самостоятельное изучение	Кол-во часов	Формы работы
			нормативных документов
Визуализация данных. Представление результатов	Подготовка отчета по лабораторной работе 6. Подготовка отчета по практической работе 1	12	Внеаудиторная, изучение учебных пособий и нормативных документов
Работа с текстовыми данными. Текстовый анализ.	Подготовка отчета по лабораторной работе 7. Подготовка к проверочному тесту	14	Внеаудиторная, изучение учебных пособий и нормативных документов

## 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

**Таблица 5 – Образовательные технологии, используемые при реализации учебных занятий**

Раздел, тема дисциплины (модуля)	Форма учебного занятия		
	Лекция	Практическое занятие, семинар	Лабораторная работа
Основы программирования на языке Python.	Обзорная лекция	Не предусмотрено	выполнение лабораторной работы
Математический аппарат.	Лекция-диалог	Не предусмотрено	выполнение лабораторной работы
Введение в модуль NumPy.	Лекция	Не предусмотрено	выполнение контрольной работы
Основы работы с Pandas.	Лекция	Не предусмотрено	выполнение лабораторной работы
Разведывательный анализ данных.	Лекция	Не предусмотрено	выполнение лабораторной работы
Визуализация данных. Представление результатов	Лекция-диалог	Не предусмотрено	выполнение лабораторной работы
Работа с текстовыми данными. Текстовый анализ.	Обзорная лекция	Не предусмотрено	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.

В рамках проведения учебного процесса используются следующие формы проведения учебных занятий:

- Интерактивные лекции,
- Тренинги,
- Обучение действием («action learning»),
- Мозговой штурм.

## **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. Перечень программного обеспечения и информационных справочных систем**

### **6.3.1. Программное обеспечение:**

Наименование программного обеспечения	Назначение
AdobeReader	Программа для просмотра электронных документов
Платформа дистанционного обучения LMS Moodle	Виртуальная обучающая среда
Mozilla FireFox	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

Среда разработки Anaconda3 (64 –bit) (в свободном доступе)

### **6.3.2. Современные профессиональные базы данных и информационные справочные системы:**

1. Электронный каталог Научной библиотеки АГУ на базе MARKSQL НПО «Информ-систем»: <https://library.asu.edu.ru>.
2. Электронный каталог «Научные журналы АГУ»: <http://journal.asu.edu.ru/>.
3. Электронно-библиотечная система elibrary. <http://elibrary.ru>
4. Справочная правовая система КонсультантПлюс: <http://www.consultant.ru>
5. 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 (модулю) ««Анализ данных в информационной безопасности»» 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 (модуля)	Код контролируемой компетенции (компетенций)	Наименование оценочного средства
1	Основы программирования на языке Python.	ПК 2, ПК 3.	Лабораторная работа №1
2	Математический аппарат.	ПК 2, ПК 3.	Лабораторная работа №2
3	Введение в модуль NumPy.	ПК 2, ПК 3.	Лабораторная работа №3
4	Основы работы с Pandas.	ПК 2, ПК 3.	Лабораторная работа №4
5	Разведывательный анализ данных.	ПК 2, ПК 3.	Лабораторная работа №5
6	Визуализация данных. Представление результатов	ПК 2, ПК 3.	Лабораторная работа №6. Практическая работа № 1
7	Работа с текстовыми данными. Текстовый анализ.	ПК 2, ПК 3.	Лабораторная работа №7. Проверочный тест. Вопросы к экзамену

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

**Таблица 7 – Показатели оценивания результатов обучения в виде знаний**

Шкала	Критерии оценивания
-------	---------------------

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 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 программирования на языке Python**

Лабораторная работа 1. Основы программирования на языке Python: типы данных и методы работы с ними (переменные, листы, словари, кортежи).

Цель: ознакомиться с основами языка Python, получить умения для выполнения дальнейших лабораторных работ.

Задачи:

- изучить типизацию данных;
- рассмотреть «ветвление» в Python;
- отработать задачи с использованием конструкции «try-except»;

В ней обучающиеся поэтапно проходят все основные аспекта языка, такие как:

1. Типизация данных.

2. Ветвления.
3. Исключения.
4. Переменные, листы, словари, кортежи

Задания для лабораторной работы № 1:

1. Опишите отличия массивов, кортежей, списков и словарей.
2. Приведите пример кода, который записывает/создает текстовый файл, записывает в него две строчки «Hello» и «123», а затем считывает его и выводит его содержимое. Код прокомментируйте.
3. Что такое исключения (ошибки) и как их можно обработать? Что лучше выбрать – написать программу так, чтобы не возникало ошибок, или чтобы ошибки были обработанными?

## Раздел 2. Математический аппарат

Лабораторная работа 2. Основы программирования на языке Python: циклы, функции, синтаксис языка.

Цель: ознакомиться с основами языка Python: циклы функции, знание синтаксиса языка.

Задачи:

- научиться пользоваться циклами «for» и «while»;
- разобрать функции и пространства имён.

В ней обучающиеся поэтапно проходят все основные аспекта языка, такие как:

1. Пространство имён.
2. Функции.
3. Циклы.

Задания для лабораторной работы № 2:

1. Приведите пример кода, реализующего проверку есть ли словаре запись с каким-то определенным ключом. Например, что контакт есть в телефонной книге.
2. При помощи цикла for выведете таблицу умножения для числа 3. Т.е. число 3 должно умножаться на каждое из [0;9] чисел и результат – выводится пользователю.

## Раздел 3. Введение в модуль NumPy

Лабораторная работа 3. Библиотека NumPy.

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

Задачи:

1. Объект ndarray.
2. Массивы в NumPy.
3. Базовые операции в NumPy.
4. Манипуляции с формой в NumPy.
5. Копии и представления в NumPy.
6. Сохранение массива в файл и чтение из файла в NumPy.

Задания для лабораторной работы № 3:

1. Перечислите наиболее важные атрибуты объектов ndarray.
2. Базовая функциональность NumPy – опишите основные доступные функции.

3. Создание массива, заполненного нулями – приведите пример кода.
4. Математические операции между массивами. Представьте код, где массив со значениями от 1 до 9 умножается на константу 2. Должна получиться таблица умножения для числа 2.
5. Слайсы (обрезка массива) – предоставите код для указанных преподавателем случаев
6. Приведите пример объединения массивов из NumPy.
7. Опишите применение унарных операций к массивам, приведите примеры.
8. Копии и представления при работе с массивами.
9. Сохранение массива в файл и чтение из файла.

#### **Раздел 4. Основы работы с Pandas**

Лабораторная работа № 4. Библиотека pandas.

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

Задачи:

1. Структура данных Series.
2. Структура данных DataFrame.
3. Доступ к данным в структурах pandas.
4. Добавление элементов в структуры pandas.
5. Обработка отсутствующих данных в pandas.
6. Сохранение объектов в файл и чтение из файла в pandas.

Задания для лабораторной работы № 4:

1. Pandas – основные сведения. Сравнение с функциональностью NumPy.
2. Опишите Конструктор класса Series.
3. Приведите пример Series и попробуйте: выводить на экран, добавлять элементы, изменять значения.
4. Опишите Конструктор класса DataFrame.
5. Повторите пункт 3 для объекта DataFrame: выводить на экран, добавлять элементы, изменять значения.
6. Открыть при помощи pandas созданный в процессе выполнения работы dataframe.csv файл с пропусками или некорректными значениями и исправить их.

#### **Раздел 5. Основные этапы анализа данных**

Лабораторная работа № 5. Библиотека scikit-learn. Обучение с учителем и без учителя.

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:

1. **Линейные:** модели, задача которых построить разделяющую или аппроксимирующую гиперплоскость (для классификации и регрессии соответственно).

2. **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).
3. **Деревья решений:** обучение моделей, базирующихся на множестве условий, оптимально выбранных для решения задачи.
4. **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).
5. **Нейронные сети:** комплексный нелинейный метод для задач регрессии и классификации.
6. **Метод опорных векторов** (англ. support vector machine, SVM): нелинейный метод, который обучается определять границы принятия решений.
7. **Наивный Байес:** прямое вероятностное моделирование для задач классификации.
8. **Метод главных компонент** (англ. principal component analysis, PCA): линейный метод понижения размерности и отбора признаков.
9. **Стохастическое вложение соседей с t–распределением** (англ. t-distributed Stochastic Neighbor Embedding, t-SNE): нелинейный метод понижения размерности.
10. **К–средних:** самый распространенный метод для кластеризации, требующий на вход число кластеров, по которым должны быть распределены данные.
11. **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.
12. **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. Классификация – что такое и где применяется.
2. Кластерный анализ – что такое и где применяется, чем отличается от классификации.
3. Линейная регрессия – проведите эксперимент с собственным датасетом (к примеру, подготовьте в Excel датасет случайных значений, лежащих возле какого–либо уравнения).
4. Проведите и проанализируйте собственные эксперименты с К–средних и деревом решений.

## Лабораторная работа № 6. Работа с изображениями и создание различных видов визуализаций

### Задание 1. Библиотека TensorFlow Keras. Классификация изображений

Цель: Написание программы для распознавания цифр на базе TensorFlow Keras. Создание нейросети и её обучение по базе данных MNIST. Тестирование MNIST выборкой и пользовательскими картинками.

Создается нейронная сеть, которая классифицирует изображения рукописных цифр. При написании программы используется надстройка «tf.keras», которая является высокоуровневой API для построения и обучения моделей в TensorFlow.

Задачи:

1. Изучение работы с датасетом MNIST.
2. Создание обучающих и тренировочных выборок.
3. Построение модели нейронной сети.
4. Обучение нейронной сети.
5. Проверка точности обученной модели.
6. Предсказание изображений с помощью обученной модели.

### Задание 2. Создание различных видов визуализаций на синтетических данных и тренировочных наборах данных

Цель: научить обучающихся основам работы с машинным зрением и показать основные алгоритмы работы с ним.

В последние годы машинное зрение получило большую огласку и вызвало интерес со стороны не только ученых, но и различных инженеров и разработчиков «интеллектуальных» приложений. В лабораторной работе обучающиеся рассмотрят вариант библиотеки OpenCV, написанной на языке «C++» для Python, опробуют некоторый её функционал и протестируют классификатор для распознавания лиц.

Задачи:

1. разобрать импорт и просмотр изображения;
2. разобрать кадрирование;
3. научиться изменять размер изображения;
4. научиться переворачивать изображение;
5. рассмотреть способ преобразование изображения в черно-белое;
6. научиться работать со сглаживанием и размытием;
7. изучить метод распознавания лиц.

### Задания для лабораторной работы № 6:

1. Что такое набор данных, или датасет? Для чего может использоваться, как может задаваться (в виде каких типов данных)?
2. Для чего необходимо разбивать датасет на обучающие выборки и валидирующие (тестируемые)? Что дает тестирование?
3. Что такое классы объектов? Зачем им нужны имена? Влияет ли имя класса на обучение, или это дополнительное «удобство» при использовании обученной нейросети?
4. Опишите назначения всех узлов элементов нейросети: входы, синапсы, нейроны, аксоны. Каким узлам соответствуют данные понятия: веса, функции активации?
5. Каким набором вызова функций библиотеки Keras можно определять слои и их параметры (функции активации нейронов на слоях, количество нейронов, тип слоя и прочие параметры)?

6. Подготовьте собственную картинку цифры. Протестируйте на ней классификацию при помощи нейросети.

### Практическая работа № 1. Создание интерактивных визуализаций и отчетов с помощью инструмента Plotly

Цель: Получение представления о функциональности, доступных методах и объектах библиотеки Plotly.

Введение в визуализацию данных. Нюансы визуализации данных и принципы человеческого восприятия. Правила создания хороших визуализаций.

Задачи:

1. Рассмотреть особенности подготовки данных для визуализации.
2. Изучить нюансы визуализации данных и принципы человеческого восприятия.
3. Научиться создавать интерактивные визуализации и отчеты с помощью инструмента Plotly.

## Раздел 7. Работа с текстовыми данными. Текстовый анализ

### Лабораторная работа № 7. Работа с текстовыми данными Задание 1. Модель, классифицирующая отзывы покупателей

Цель: Построение прогностической модели, анализирующей отзывы покупателей о продукции и сортирующей их по двум классам: положительные отзывы и отрицательные. Модель учится на уже классифицированных образцах предсказывать класс других образцов и реализована с помощью scikit-learn.

Задачи:

1. Получение набора образцов.
2. Преобразование текста в числовые векторы признаков.
3. Обучение и оценка модели.

### Задание 2. Обработка естественного языка

Цель: Научиться предварительно обрабатывать и анализировать содержание текстов, используя библиотеку nltk и метод мешка слов, а также осуществлять перевод текста с помощью библиотеки googletrans.

Задачи:

1. Предварительная оценка текста.
2. Разделение на предложения.
3. Разделение на слова.
4. Перевод в нижний регистр, удаление стоп-слов и знаков пунктуации.
5. Лемматизация.
6. Стемминг.
7. Мешок слов.
8. Определение тональности текста.
9. Перевод и величины достоверности для определения языка.

Задания для лабораторной работы № 7:

1. С какой целью используется разделение текста на предложения и слова? Как называется этот процесс? Приведите пример.
2. С какой целью используется перевод в нижний регистр, удаление стоп-слов и знаков пунктуации? Приведите пример.
3. С какой целью используется лемматизация? Приведите пример.
4. С какой целью используется стемминг? Приведите пример.

5. Опишите принцип метода мешка слов. Приведите пример. С какой целью используется этот метод.
6. Как осуществляется определение тональности текста с помощью библиотеки nltk? Приведите примеры.
7. Как осуществляется перевод текста? Приведите примеры.
8. Какие коллизии могут возникать при переводе текстов? Приведите примеры.
9. Как определить величину достоверности для определения языка с которого осуществляется перевод? Приведите примеры.
10. Как определить язык исходного текста? Приведите примеры.
11. Выполнить загрузку из указанного файла так, как это указано в строке кода ниже и осуществить обработку текста.

### Проверочный тест

Какой метод используется для предсказания категориальных переменных?

- 1) Линейная регрессия
- 2) Логистическая регрессия
- 3) Кластеризация
- 4) ARIMA

Что такое нормализация данных?

- 1) Процесс удаления выбросов
- 2) Процесс приведения данных к единой шкале
- 3) Процесс кластеризации данных
- 4) Процесс классификации данных

Какой метод используется для предсказания числовых значений?

- 1) Линейная регрессия
- 2) Логистическая регрессия
- 3) Кластеризация
- 4) ARIMA

Выберите верное утверждение:

- 1) Чем ближе к 1 индекс корреляции, тем выше качество модели множественной регрессии.
- 2) Чем ближе к 0 коэффициент детерминации, тем выше качество модели множественной регрессии.
- 3) Независимость остатков проверяется с помощью критерия Дарбина – Уотсона.
- 4) Качество регрессора характеризуется фактом обоснованной зависимости остатков от целевой переменной.

Поставьте в соответствие библиотеку и ее описание

- 1) Scikit-learn
  - 2) NumPy
  - 3) Pandas
  - 4) Seaborn
- a) библиотека для работы с многомерными массивами числовых данных и со сложными математическими операциями.
  - b) библиотека визуализации, основанная на Matplotlib.
  - c) опенсорсная библиотека машинного обучения Python, с широким спектром алгоритмов

кластеризации, регрессии и классификации.

- d) это библиотека с открытым исходным кодом, которая предлагает широкий спектр инструментов для обработки и анализа данных. С ее помощью можно читать данные из широкого спектра источников, таких как CSV, базы данных SQL, файлы JSON и Excel.

### Вопросы к экзамену:

1. Язык Python и особенности его стиля программирования. Интерактивный режим Python. IPython. Jupyter Notebook.
2. Синтаксис и управляющие конструкции языка Python. Переменные, значения и их типы. Типы данных в Python.
3. Встроенные операции и функции. Основные алгоритмические конструкции.
4. Условный оператор. Множественное ветвление.
5. Циклы и счетчики.
6. Определение функций. Параметры и аргументы. Вызовы функций. Оператор возврата. Конструкции \*args, \*\*kwargs.
7. Списки, кортежи и словари.
8. Операторы общие для всех типов последовательностей.
9. Специальные операторы и функции для работы со списками. Срезы.
10. Работа со словарями. Методы словарей.
11. Случайные числа. random, randrange, choice.
12. Функции обработки строк. join, replace, split.
13. Стандартная библиотека и pip. Модули и пакеты в Python. Основные стандартные модули.
14. Импортирование модулей. Создание собственных модулей и их импортирование. Специализированные модули и приложения.
15. Файлы и исключения. Работа с внешними источниками данных.
16. Исключения, обработка исключений, вызов исключений (try-except-finally).
17. Утверждения (assert). Открытие, чтение, запись. (open, инструкция with).
18. Работа с текстовыми файлами, xml и csv - файлами.
19. Функциональное программирование. Лямбда-функции.
20. Использование функций map, filter, reduce, zip.
21. Генераторы, декораторы, рекурсия.
22. Модификация функций с помощью декораторов.
23. Итерируемые объекты. Использование генераторов (yield).
24. ООП в Python. Классы, объекты и экземпляры классов. Наследование.
25. Магические методы. Переопределение операторов. Методы классов.
26. Инкапсуляция. Условно частные и строго частные методы.
27. Регулярные выражения. Использование регулярных выражений. Пакет re.
28. Наука о данных и Python. Библиотеки: NumPy, pandas, matplotlib, SciPy.
29. Основы NumPy: массивы и векторные вычисления.
30. Инструменты визуализации данных для Python.
31. Введение в API библиотеки matplotlib.
32. Библиотека pandas. Введение в структуры данных pandas.
33. Объекты Dataframe и Series.
34. Визуализация данных в pandas. Seaborn.
35. Агрегирование данных и групповые операции.
36. Сбор и подготовка данных в Python: извлечение данных с web-страниц (web-scraping). Библиотека BeautifulSoup.
37. Работа с динамическими сайтами с помощью Selenium.
38. Массовый скрепинг с помощью scrapy.
39. Работа со структурированными данными: JSON и XML.

#### 40. Открытые API. Telegram API, VK API.

**Таблица 9. Примеры оценочных средств с ключами правильных ответов**

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
ПК-2. Способен выполнять работы по установке, настройке и техническому обслуживанию защищенных технических средств обработки информации.				
1)	Задание закрытого типа	Какой метод используется для предсказания категориальных переменных?  1) Линейная регрессия 2) Логистическая регрессия 3) Кластеризация 4) ARIMA	3	2
2)		Что такое нормализация данных?  1) Процесс удаления выбросов 2) Процесс приведения данных к единой шкале 3) Процесс кластеризации данных 4) Процесс классификации данных	2	2
3)		Какой метод используется для предсказания числовых значений?  1) Линейная регрессия 2) Логистическая регрессия 3) Кластеризация 4) ARIMA	1	2
4)		Как связаны между собой AUC и ROC 1. ROC-«кривая ошибок», а AUC – площадь под ней 2. ROC- «кривая правильных ответов» , а AUC- площадь над ней 3. ROC – график, а AUC-оптимальное пороговое значение 4. AUC-«кривая ошибок», а ROC- площадь под ней 5. AUC - «кривая правильных ответов», а ROC - площадь над ней 6. AUC – график, а ROC -оптимальное пороговое значение	1	3
5)		Выберите верное утверждение: 1. Чем ближе к 1 индекс корреляции, тем выше качество модели множественной регрессии. 2. Чем ближе к 0 коэффициент детерминации, тем выше качество модели множественной регрессии. 3. Независимость остатков проверяется с помощью критерия Дарбина – Уотсона. 4. Качество регрессора характеризуется фактом обоснованной зависимости остатков от целевой переменной.	1, 3	3
6)		Задание	Что такое точность классификации?	Точность-это доля правильно

	открыто о типа		распознанных экземпляров	
7)		Какие признаки называются категориальными?	Признаки называются категориальными, если их можно отнести к какой-либо группе, но при этом не важен порядок	6
8)		В чем выражается проблема переобучения?	Модель эффективно работает только с теми данными, на которых была обучена	8
9)		Дайте определение понятию «классификация».	Классификация – это задача отнесения объекта к какому-либо из заранее определенных классов. Относится к типу «обучение с учителем».	8
10)	Задание комбинированного типа	Функция библиотеки Scikit-learn, которая используется для подсчета средней абсолютной ошибки регрессора?  а) mean_absolute_error  б) predict()  в) train_test_split  г) accuracy_score  д) roc_curve	Функция библиотеки Scikit-learn mean_absolute_error применяется для подсчета средней абсолютной ошибки регрессора.	8
ПК-3. Способен осуществлять внедрение систем защиты информации для обеспечения информационной безопасности автоматизированных систем.				
11)	Задания закрытого типа	Поставьте в соответствие библиотеку и ее описание 5) Scikit-learn 6) NumPy 7) Pandas 8) Seaborn е) библиотека для работы с многомерными массивами числовых данных и со сложными математическими операциями. ф) библиотека визуализации, основанная на Matplotlib. г) опенсорсная библиотека машинного обучения Python, с широким спектром алгоритмов кластеризации, регрессии и классификации. h) это библиотека с открытым исходным кодом, которая предлагает широкий	1-с 2-а 3-д 4-б	2

		спектр инструментов для обработки и анализа данных. С ее помощью можно читать данные из широкого спектра источников, таких как CSV, базы данных SQL, файлы JSON и Excel.		
12)		С какой целью используется класс OneHotEncoder() библиотеки Scikit Learn? 1) Преобразование категориальных признаков в массив чисел. 2) Уменьшение размерности 3) Кодирование категориальных признаков в UTF-8 4) Преобразование количественных признаков в структурированные	1	2
13)		Под доставкой модели в машинном обучении понимается: 1) Отправка модели заказчику 2) Сборка, тестирование и запуск модели в виде ПО с внесенными изменениями 3) Публикация программного кода модели на GitHub 4) Запуск и развертывание модели в виде ПО в фактический продукт	4	2
14)		Продолжите фразу: "Команда Data Science ...": 15) Отвечает за получение и хранение данных заказчика 16) Разрабатывает пользовательский интерфейс для работы с данными 17) Разрабатывает математическое и алгоритмическое решение проблемы 18) Реализует передачу исходных данных разработчикам	3	2
15)		DevOps сконцентрирован на... 1. быстрой и непрерывной поставке работающего программного обеспечения 2. обработке данных, а также возможностях оперативного и безопасного доступа к ним 3. на создании инфраструктуры для корректного хранения, движения и использования данных	1	2
16)		Опишите функцию библиотеки Scikit-learn, которая используется для разделения множества объектов на обучающее и проверочное множества?	train_test_split. Используется для разделения набора данных на подмножества, чтобы свести к минимуму возможность смещения в процессе оценки и проверки.	8
17)		Опишите функцию библиотеки Scikit-learn, которая используется для определения верности классификации?	accuracy_score. Применяется для оценки качества	8

			модели.	
18)		Опишите функцию библиотеки Scikit-learn, которая используется для анализа ROC-кривых?	roc_curve. Используется для возврата кривой ROC данной модели.	8
19)		Назначение библиотеки pandas	Библиотека pandas предоставляет специальные структуры данных и операции для манипулирования числовыми таблицами и временными рядами: объекты <i>DataFrame</i> и <i>Series</i> .	2
20)	Задание комбинированного типа	Опишите функцию библиотеки Scikit-learn, которая используется для разделения множества объектов на обучающее и проверочное множества?  а) mean_absolute_error б) predict() в) train_test_split г) accuracy_score д) roc_curve	Функция библиотеки Scikit-learn train_test_split. Используется для разделения набора данных на подмножества, чтобы свести к минимуму возможность смещения в процессе оценки и проверки.	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)

##### Экзамен

Экзаменационный билет включает в себя 2 теоретических вопроса и 1 задачу.

Основаниями для снижения оценки являются:

- ошибки в объяснениях и комментариях при верно выполненном задании;
- неполный ответ;
- наличие мелких неточностей или незначительных искажений фактов;
- передача экзамена (1я – минус 5 баллов, 2я и последующие – минус 10 баллов).

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

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

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

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

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

№ п/п	Контролируемые мероприятия	Количество мероприятий / баллы	Максимальное количество баллов	Срок представления
<b>Основной блок</b>				
1.	<i>Выполнение лабораторной работы</i>	7/5	35	В соответствии с таблицей 2
2.	<i>Выполнение практической работы</i>	1/4	4	
3.	<i>Тест</i>	1/1	1	
<b>Всего</b>			<b>40</b>	-
<b>Блок бонусов</b>				
4.	<i>Посещение занятий без пропусков</i>		3	
5.	<i>Своевременное выполнение всех заданий</i>		3	
6.	<i>Активность студента на занятии</i>		4	
<b>Всего</b>			<b>10</b>	-
<b>Дополнительный блок</b>				
7.	<i>Экзамен</i>		50	
<b>Всего</b>			<b>50</b>	-
<b>ИТОГО</b>			<b>100</b>	-

**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)
85–89	4 (хорошо)
75–84	
70–74	
65–69	3 (удовлетворительно)
60–64	
Ниже 60	2 (неудовлетворительно)

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. Кузьмич, Р. И. Модификации метода логического анализа данных для задач классификации : монография / Р. И. Кузьмич, И. С. Масич. — Красноярск : Сибирский федеральный университет, 2018. — 181 с. — ISBN 978-5-7638-3698-1. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/84252.html>

2. Воронова, Л. И. Machine Learning: регрессионные методы интеллектуального анализа данных : учебное пособие / Л. И. Воронова, В. И. Воронов. — Москва : Московский технический университет связи и информатики, 2018. — 82 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/81325.html>

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>

### **8.2. Дополнительная литература:**

1. 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>

2. 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>

3. 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](http://www.studentlibrary.ru).

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

## **9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)**

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).