

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

_____ И.М. Ажмухамедов

«06» июня 2024 г.

УТВЕРЖДАЮ

И.о. заведующего кафедрой ИБ

_____ Т.Г. Гурская

«06» июня 2024 г.

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

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»
бакалавр**

Квалификация (степень)

Форма обучения

очная

Год приема

2021

Курс

4

Семестр

7

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1. Целями освоения дисциплины (модуля) «Анализ данных в информационной безопасности» являются освоение и развитие навыков программирования на языке Python, представления о сборе, обработке и анализе данных, 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.06.02 «Анализ данных в информационной безопасности» относится к части, формируемая участниками образовательных отношений и осваивается в 7 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, основ программирования на языке Python.

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

- Математические основы защиты информации
- Основы программирования.

знания:

- роль и значение информационных ресурсов в современном обществе,
- виды и формы информации,
- современные информационные технологии обработки информации,
- этапы и методы ее обработки информации,

умения:

- применять компьютерную технику и информационные технологии для обработки информации, и решения практических задач,

навыки:

- владения инструментальными средствами информационных технологий обработки информации,
- владения инфокоммуникационных технологий.

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):

- способности осуществлять внедрение систем защиты информации для обеспечения информационной безопасности автоматизированных систем (ПК-3);

- способности администрировать средства защиты информации в компьютерных системах и сетях (ПК-4).

Таблица 1 – Декомпозиция результатов обучения

Код компетенции и наименование	Планируемые результаты обучения по дисциплине (модулю)		
	Знать	Уметь	Владеть
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 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	ИПК-3.2. Уметь: администрировать программные средства защиты информации автоматизированных систем, устранять известные уязвимости автоматизированной системы, 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	ИПК-3.3. Владеть: методикой анализа структурных и функциональных схем защищенной автоматизированной системы
ПК-4. Способен администрировать средства защиты информации в компьютерных системах и сетях	ИПК-4.1. Знать: источники угроз информационной безопасности в компьютерных сетях и меры по их предотвращению; принципы функционирования программных средств криптографической защиты информации; виды политик управления доступом и информационными потоками в компьютерных сетях;	ИПК-4.2. Уметь: анализировать угрозы безопасности информации в компьютерных системах и сетях; настраивать правила обработки пакетов в компьютерных сетях; настраивать политики безопасности операционных систем, оценивать угрозы безопасности информации в компьютерных системах и сетях;	ИПК-4.3. Владеть: навыками управления средствами межсетевого экранирования в компьютерных сетях методикой оценки оптимальности выбора программно-аппаратных средств защиты информации и их режимов функционирования в операционных системах

	требования по составу и характеристикам подсистем защиты информации применительно к операционным системам; принципы работы и правила эксплуатации программно-аппаратных средств защиты информации	противодействовать угрозам безопасности информации с использованием встроенных средств защиты информации операционных систем, настраивать антивирусные средства защиты информации в операционных системах	
--	---	---	--

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Объем дисциплины (модуля) в зачетных единицах (3 зачетные единицы) 108 часов, на контактную работу обучающихся с преподавателем выделено 54 часа (из них лекции – 18 часов, лабораторных работ – 36 часа) и на самостоятельную работу обучающихся 54 часа.

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 (refresher).	7	1-2	2		4		8	Лабораторная работа №1
2	Математический аппарат (refresher).	7	3-4	2		4		8	Лабораторная работа №2
3	Введение в модуль NumPy.	7	5-6	2		4		8	Лабораторная работа №3
4	Основы работы с Pandas.	7	7-8	2		4		8	Лабораторная работа №4
5	Разведывательный анализ данных.	7	9-10	2		4		8	Лабораторная работа №5
6	Визуализация данных. Представление результатов	7	11-14	4		8		6	Лабораторная работа №6
7	Работа с текстовыми данными. Текстовый анализ.	7	15-18	4		8		8	Лабораторная работа №7
ИТОГО			108	18		36		54	ЭКЗАМЕН

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-3	ПК-4	
Основы программирования на языке Python (refresher).	14	+	+	2
Математический аппарат (refresher).	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
Итого	108			2

СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Лекция 1. Основы программирования на языке Python (refresher).

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: типы данных и методы работы с ними (переменные, листы, словари, кортежи).

Лекция 2. Математический аппарат (refresher).

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

Лекция 3. Введение в модуль NumPy.

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

Лекция 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.

Лекция 5. Разведывательный анализ данных.

Регрессионные модели. Метод наименьших квадратов. Логистическая регрессия. Проведение разведывательного анализа данных в Pandas на примере.

Лекция 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.

Лекция 7. Работа с текстовыми данными. Текстовый анализ.

Введение в анализ текста. Применение в политологии. Особенности подготовки данных
Создание интерактивных визуализаций и отчетов с помощью инструмента Plotly.

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 (refresher).	Подготовка отчета по лабораторной работе 1	8	Внеаудиторная, изучение учебных пособий и нормативных документов
Математический аппарат (refresher).	Подготовка отчета по лабораторной работе 2	8	Внеаудиторная, изучение учебных пособий и нормативных документов
Введение в модуль NumPy.	Подготовка отчета по лабораторной работе 3	8	Внеаудиторная, изучение учебных пособий и нормативных документов
Основы работы с Pandas.	Подготовка отчета по лабораторной работе 4	8	Внеаудиторная, изучение учебных пособий и нормативных документов
Разведывательный анализ данных.	Подготовка отчета по лабораторной работе 5	8	Внеаудиторная, изучение учебных пособий и нормативных документов
Визуализация данных.	Подготовка отчета по лабораторной работе 6	6	Внеаудиторная, изучение учебных

Номер раздела (темы)	Темы/вопросы, выносимые на самостоятельное изучение	Кол-во часов	Формы работы
Представление результатов			пособий и нормативных документов
Работа с текстовыми данными. Текстовый анализ.	Подготовка отчета по лабораторной работе 7	8	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 (refresher).	Обзорная лекция	Не предусмотрено	выполнение лабораторной работы
Математический аппарат (refresher).	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. Программное обеспечение, современные профессиональные базы данных и информационные справочные системы

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 (refresher).	ПК 3, ПК 4.	Лабораторная работа №1
2	Математический аппарат (refresher).	ПК 3, ПК 4.	Лабораторная работа №2
3	Введение в модуль NumPy.	ПК 3, ПК 4.	Лабораторная работа №3
4	Основы работы с Pandas.	ПК 3, ПК 4.	Лабораторная работа №4
5	Разведывательный анализ данных.	ПК 3, ПК 4.	Лабораторная работа №5
6	Визуализация данных. Представление результатов	ПК 3, ПК 4.	Лабораторная работа №6
7	Работа с текстовыми данными. Текстовый анализ.	ПК 3, ПК 4.	Лабораторная работа №7. Вопросы к экзамену

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

Таблица 7 – Показатели оценивания результатов обучения в виде знаний

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует глубокое знание теоретического материала, умение обоснованно излагать свои мысли по обсуждаемым вопросам, способность полно, правильно и аргументированно отвечать на вопросы, приводить примеры
4	демонстрирует знание теоретического материала, его последовательное

«хорошо»	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)

Основы программирования на языке Python (refresher).

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

Математический аппарат (refresher).

Лабораторная работа 2. Задание 1. Основы программирования на языке Python: циклы функции, знание синтаксиса 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.

Задание 2. Классификация данных

Цель: научиться работать с данными при помощи визуальных инструментов и разобрать азы классификации при помощи построения простейшего классификатора со статичными параметрами.

Задачи:

- научиться анализировать данные;
- сформировать понятие математических срезов;
- получить умения в работе с визуальными инструментами;
- построить классификатор на основе данных полученных при анализе;
- научиться калибровать нейросеть для получения более точных ответов.

В этой лабораторной работе начинается изучение принципов машинного обучения. В ней обучающиеся будут учиться работать с графической информацией и производить анализ полученных данных и самостоятельно строить простейший классификатор. Данная работа поможет понять простейшие понятия и принципы работы более сложных классификаторов.

Введение в модуль NumPy.

Лабораторная работа 3. Задание 1. Особенности типы данных в NumPy. Работа с векторами и матрицами. Вычисление главных статистических метрик с помощью NumPy (среднее, медиана, мода, дисперсия).

Цель: ознакомиться с основами NumPy, работать с векторами и матрицами, статистическими метриками.

Задачи:

- изучить особенности типы данных в NumPy,
- научиться работать с векторами и матрицами.
- получить умения вычислять главные статистические метрики с помощью NumPy (среднее, медиана, мода, дисперсия).

Задание 2. Классификация методом "К-ближайших соседей"

Цель: изучить метод простейший метод классификации данных «К-ближайших соседей» и научиться производить оценку данных с помощью визуальных инструментов Python.

Задачи:

- детально разобрать метод машинного обучения «К-ближайших соседей»;
- научиться работать с информацией;
- сформировать понятие математических срезов;
- получить умения в работе с визуальными инструментами;
- научиться калибровать нейросеть для получения более точных ответов.

В этой лабораторной работе будет рассматриваться метод машинного обучения «К-ближайших соседей», а также его положительные и отрицательные стороны; рассмотрим библиотеку для визуализации данных и математического анализа «matplotlib»; будем производить калибровку нейронной сети для получения максимального процента правильных ответов, а также обсудим проблему переобучения в нейросетях.

Основы работы с Pandas.

Лабораторная работа № 4. Работа с табличными данными в Pandas на примере.

Цель: научиться пользоваться библиотекой Pandas и её встроенными объектами для визуализации данных в датасетах.

Задачи:

- получить умения по использованию библиотеки Pandas;
- сформировать понятия о DataFrame и Series;
- научиться строить графики с помощью scatter matrix (матрица рассеивания) и matplotlib.

Pandas — это библиотека языка Python, которая позволяет создавать объекты, в которых данные хранятся в табличной форме. Это основной структурный объект необходимый для анализа и очистки данных перед их использованием в обучении модели. Это первая лабораторная работа, связанная с Pandas. В ней обучающиеся разберут структуру и основные функции данной библиотеки.

Разведывательный анализ данных.

Лабораторная работа № 5. Проведение разведывательного анализа данных в Pandas на примере.

Задание 1. Анализ данных с помощью Pandas

Цель: научиться пользоваться библиотекой Pandas и её встроенными объектами для анализа данных в датасетах.

Задачи:

- получить умения по использованию библиотеки Pandas;
- научиться анализировать и обрабатывать данные с помощью Pandas;
- закрепить умения визуализации в Pandas.

Обучающиеся переходят к анализу данных с помощью библиотеки Pandas. В данной работе обучающиеся научатся делать срезы, группировки, индексацию данных по разным параметрам, а также строить графики по созданным срезам. Знания и умения, полученные в ходе выполнения работы помогут понять, некоторые механизмы машинного обучения.

Задание 2. Линейная регрессия

Цель: понять и научиться применять метод линейной регрессии в машинном обучении.

Задачи:

- изучить модель линейного регрессора;
- произвести обучение модели;
- рассмотреть особенности данного метода машинного обучения;
- произвести предсказание на основе созданной модели.

Линейная регрессия является одним из самых универсальных методов машинного обучения, т.к. может принимать на вход данные различных типов и структур. Также данная модель достаточно устойчива к переобучению и имеет большую скорость обработки данных. В этой лабораторной будут рассматриваться все нюансы и принципы работы этого метода классификации.

Визуализация данных. Представление результатов

Лабораторная работа № 6. 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.

Задачи:

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

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, опробуем некоторый её функционал и протестируем классификатор для распознавания лиц.

Работа с текстовыми данными. Текстовый анализ.

Лабораторная работа № 7. Задание 1. Особенности подготовки данных. Создание интерактивных визуализаций и отчетов с помощью инструмента Plotly.

Цель: научиться работать с текстовыми данными.

Задачи:

- рассмотреть особенности подготовки данных,
- научиться создавать интерактивные визуализации и отчеты с помощью инструмента Plotly

Задание 2. Деревья решений

Цель: познакомить обучающихся с методом машинного обучения, построенном на деревьях решений, а также научить строить сами деревья.

Задачи:

- рассмотреть понятие дерева решений;
- рассмотреть варианты применения данной классификации;
- обучить модель на основе классов;
- отобразить дополнительный класс на модели и посмотреть результат;
- рассмотреть плюсы и минусы данной модели.

Одним из самых простых для понимания является классификатор «Дерево решений». Эта модель строит «дерево» классов, у каждого из которых есть свои параметры. Попадающие под эти параметры данные причисляются к тому или иному классу. Модель очень чувствительна к переобучению и данным, которые получает на вход. Лабораторная работа поможет нам разобраться во всех тонкостях данного метода и покажет все его нюансы.

Задание 3. Метод случайного леса

Цель: сформировать понятие случайного леса, а также научить обучающихся использовать данную модель для решения задач.

Задачи:

- рассмотреть понятие случайного леса;
- рассмотреть пример кода для решения простых задач;
- научить подбирать параметры модели для улучшения качества прогнозов модели.

Случайный лес является модифицированным вариантом «Древа решений» и не имеет всех его отрицательных особенностей, например, он не чувствителен к переобучению. В лабораторной работе будет рассмотрено строение данного метода классификации и его устройство.

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-3 – способен осуществлять внедрение систем защиты информации для обеспечения информационной безопасности автоматизированных систем				
1.	Задание закрыт	На рисунке представлен график покрытия, изображающий сравнение двух моделей. Какая модель доминирует?	3	2

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
	ого типа	График покрытия 1) M1 2) M2 3) Ни одна		
2.		ROC-кривая описывает... 1) зависимость ложных срабатываний от размера обучающего множества 2) чувствительность модели к разным порогам классификации 3) качество модели при тестировании с использованием различных проверочных множеств 4) частоту истинно положительных ответов модели	2	3
3.		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
4.		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
5.		Для решения каких задач информационной безопасности обычно используются методы машинного обучения? 1) Распознавание лиц 2) Составление модели нарушителя	1,4	3

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
		3) Прогнозирование числа кибер-атак 4) Обнаружение аномалий в сетевом трафике 5) Построение модели угроз		
6.	Задание открытого типа	Что такое точность классификации?	Точность-это доля правильно распознанных экземпляров.	5
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.		В проверочном множестве 800 объектов: 300 объектов класса <i>a</i> , 500 объектов класса <i>b</i> . Правильно были распознаны 275 объектов класса <i>a</i> и 480 объектов класса <i>b</i> . Рассчитайте верность (ассигасу) классификатора.	0,94	8
10.		Дайте определение понятию «классификация».	Классификация – это задача отнесения объекта к какому-либо из заранее определенных классов. Относится к типу «обучение с учителем».	8
ПК-4 – способен администрировать средства защиты информации в компьютерных системах и сетях				
11.	Задание закрытого типа	Поставьте в соответствие библиотеку и ее описание 1. Scikit-learn 2. NumPy 3. Pandas 4. Seaborn а. библиотека для работы с многомерными массивами числовых данных и со сложными математическими операциями. б. библиотека визуализации, основанная на Matplotlib. с. опенсорсная библиотека машинного обучения Python, с широким спектром алгоритмов кластеризации, регрессии и классификации. д. 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.	1-с 2-а 3-д 4-б	2
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:	4	2

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
		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		
14.		Продолжите фразу: “Команда Data Science ...”: 1. Отвечает за получение и хранение данных заказчика 2. Разрабатывает пользовательский интерфейс для работы с данными 3. Разрабатывает математическое и алгоритмическое решение проблемы 4. 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.	Задание открытого типа	Опишите функцию библиотеки Scikit-learn, которая применяется для предсказания класса?	Predict. Функция predict() используется для прогнозирования значений на основе предыдущего поведения данных и путем подгонки этих данных к модели.	8
17.		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
18.		Опишите функцию библиотеки Scikit-learn, которая используется для определения верности классификации?	accuracy_score. Применяется для оценки качества модели.	8
19.		Опишите функцию библиотеки Scikit-learn, которая используется для анализа ROC-кривых?	roc_curve. Используется для возврата кривой ROC данной модели.	8
20.		Опишите функцию библиотеки Scikit-learn, которая используется для подсчета средней абсолютной ошибки регрессора?	mean_absolute_error. Применяется для подсчета средней абсолютной ошибки регрессора.	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
Основной блок				
1.	<i>Выполнение лабораторной работы</i>	7/5	35	В соответствии с таблицей 2
2.	<i>Тест</i>	1/5	5	
Всего			40	-
Блок бонусов				
3.	<i>Посещение занятий без пропусков</i>		3	
4.	<i>Своевременное выполнение всех заданий</i>		3	
5.	<i>Активность студента на занятии</i>		4	
Всего			10	-
Дополнительный блок				
6.	<i>Экзамен</i>		50	
Всего			50	-
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

Показатель	Балл
<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.

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).