

МИНОБРНАУКИ РОССИИ
АСТРАХАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

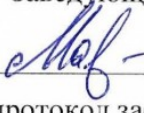
СОГЛАСОВАНО
Руководитель ОПОП

 А. Н. Марьенков

«11» мая 2021 г.

УТВЕРЖДАЮ

Заведующий кафедрой ИБиЦТ

 А. Н. Марьенков
протокол заседания кафедры №10

«11» мая 2021 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
ПРОГРАММИРОВАНИЕ ДЛЯ АНАЛИЗА ДАННЫХ

Составитель(-и)

**Проталинский И.О., к.т.н., доцент кафедры
ИБиЦТ**

Направление подготовки /
специальность

**09.03.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И
ТЕХНОЛОГИИ**

Направленность (профиль) ОПОП

**Технологии разработки и администрирования
информационных систем**

Квалификация (степень)

бакалавр

Форма обучения

очная

Год приема

2021

Курс

3

Астрахань – 2021 г.

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1. Целями освоения дисциплины (модуля) программирование для анализа данных 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) относится к вариативной части (обязательные дисциплины).

Дисциплина «Программирование для анализа данных» 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 дисциплинами (модулями):

Математические основы информационных технологий и вычислительной техники

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 инструментальных средств информационных технологий обработки информации,

владения инфокоммуникационных технологий.

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

– Информационные системы в управлении предприятием.

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

Процесс изучения дисциплины (модуля) направлен на формирование элементов

следующих компетенций в соответствии с ФГОС ВО и ОПОП ВО по данному направлению подготовки (специальности):

а) общепрофессиональных (ОПК):

ОПК–6 Способен разрабатывать алгоритмы и программы, пригодные для практического применения в области информационных систем и технологий.

б) профессиональных (ПК):

ПК–2 Способен разрабатывать программное обеспечение, включая проектирование, отладку, проверку работоспособности и модификацию ПО.

Таблица 1

Декомпозиция результатов обучения

Код компетенции	Планируемые результаты освоения дисциплины (модуля)		
	Знать	Уметь	Владеть
ОПК–6	ИОПК-6.1.1 методы алгоритмизации, языки и технологии программирования, пригодные для практического применения в области информационных систем и технологий.	ИОПК-6.2.1 применять методы алгоритмизации, языки и технологии программирования при решении профессиональных задач в области информационных систем и технологий.	ИОПК-6.3.1 программирования, отладки и тестирования прототипов программно-технических комплексов задач.
ПК–2	ИПК-2.1.1 современные информационные технологии разработки, отладки, проверки работоспособности, модификации программного обеспечения	ИПК-2.2.1 осуществлять выбор информационных технологий для решения задач по разработке, отладке, проверке работоспособности, модификации программного обеспечения	ИПК-2.3.1 навыками разработки, отладки, проверки работоспособности, модификации программного обеспечения с использованием современных информационных технологий

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Объем дисциплины (модуля) составляет 4 зачетные единицы, в том числе 68 часов, выделенных на контактную работу обучающихся с преподавателем (из них 34 часа – лекции, 34 часа – лабораторные работы) и 76 - часов на самостоятельную работу обучающихся.

Таблица 2

Структура и содержание дисциплины (модуля)

№ п/п	Наименование раздела, темы	Неделя семестр	Контактная работа (в часах)	Самостоят. работа	Формы текущего контроля успеваемости (по неделям семестра)
-------	----------------------------	----------------	-----------------------------	-------------------	--

		Семест	Р а	Л	ПЗ	ЛР	КР	СР	Форма промежуточной аттестации (по семестрам)
1	Основы программирования на языке Python (refresher).	5	1-2	5		5		10	Лабораторная работа №1
2	Математический аппарат (refresher).		3-5	5		5		10	Лабораторная работа №2
3	Введение в модуль NumPy.		6-8	5		5		10	Лабораторная работа №3
4	Основы работы с Pandas.		9-11	5		5		10	Лабораторная работа №4
5	Разведывательный анализ данных.		12-14	5		5		10	Лабораторная работа №5
6	Визуализация данных. Представление результатов		15-16	3		3		10	Лабораторная работа №6
7	Работа с текстовыми данными. Текстовый анализ.		17-18	6		6		16	Лабораторная работа №7
ИТОГО				34		34		76	ЭКЗАМЕН

Условные обозначения:

Л – занятия лекционного типа; ПЗ – практические занятия, ЛР – лабораторные работы; КР – курсовая работа; СР – самостоятельная работа по отдельным темам

Таблица 3

Матрица соотнесения разделов, тем учебной дисциплины (модуля)
и формируемых в них компетенций

Разделы, темы дисциплины (модуля)	Кол-во часов	Компетенции		общее количество компетенций
		ОПК-6	ПК-2	
Основы программирования на языке Python (refresher).	20	+	+	2
Математический аппарат (refresher).	20	+	+	2
Введение в модуль NumPy.	20	+	+	2
Основы работы с Pandas.	20	+	+	2
Разведывательный анализ данных.	20	+	+	2
Визуализация данных. Представление результатов	26	+	+	2

Работа с текстовыми данными. Текстовый анализ.	28	+	+	2
Итого	144			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. ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

5.1. Указания по организации и проведению лекционных, практических (семинарских) и лабораторных занятий с перечнем учебно-методического обеспечения

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	10	Внеаудиторная, изучение учебных пособий и нормативных документов
Математический аппарат (refresher).	Подготовка отчета по лабораторной работе 2	10	Внеаудиторная, изучение учебных пособий и нормативных документов
Введение в модуль NumPy.	Подготовка отчета по лабораторной работе 3	10	Внеаудиторная, изучение учебных пособий и нормативных документов
Основы работы с Pandas.	Подготовка отчета по лабораторной работе 4	10	Внеаудиторная, изучение учебных пособий и нормативных документов
Разведывательный анализ данных.	Подготовка отчета по лабораторной работе 5	10	Внеаудиторная, изучение учебных пособий и нормативных документов
Визуализация данных. Представление результатов	Подготовка отчета по лабораторной работе 6	10	Внеаудиторная, изучение учебных пособий и нормативных документов
Работа с текстовыми данными. Текстовый анализ.	Подготовка отчета по лабораторной работе 7	16	Внеаудиторная, изучение учебных пособий и нормативных документов

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

Отчет по лабораторным работам.

Результатом работы, выполняемой студентами, является электронный отчет по выполнению лабораторно-практической работы, тематика которых представлена в таблице 4,

Электронный отчет представляет собой файл формата Jupiter Notebook, содержащий программный код, результаты выполнения программы и текстовые пояснения. Код должен быть исполняемым и его запуск должен приводить к тем же результатам, что и в предоставленном отчете. Файл передается на проверку преподавателю путем загрузки на ресурс <http://moodle.asu.edu.ru> в соответствующий заданию раздел.

Задания к лабораторно-практическим занятиям размещены на образовательном портале <http://moodle.asu.edu.ru>. Рекомендуется заранее ознакомиться с темой, основными вопросами, рекомендациями, требованиями к представлению отчета и критериями оценивания заданий.

В процессе подготовки к лабораторно-практическим занятиям, необходимо обратить особое внимание на самостоятельное изучение рекомендованной литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной литературой, материалами периодических изданий и Интернета является наиболее эффективным методом получения дополнительных знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала.

Разработка, развитие и регулярная демонстрация мини-проекта в аудитории с обсуждением.

Студенты разрабатывают собственный классификатор изображений, который создается на первых лекциях и развивается на протяжении всего курса путем постепенного добавления приемов и методик, освоенных в процессе выполнения лабораторных работ. Студенты сравнивают свои классификаторы и обмениваются между собой опытом, развивая навыки коллективной работы над проектом. Приветствуется разделение условий эксперимента над моделями и коллективное обсуждение отличий в результатах экспериментов.

Рекомендации по работе с литературными и информационными источниками сети Интернет

Работу с информационными источниками целесообразно начать с изучения общих работ по машинному обучению, а также публикаций последних лет, в которых содержатся основные вопросы изучаемой проблемы.

6. ОБРАЗОВАТЕЛЬНЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

При реализации различных видов учебной работы по дисциплине могут использоваться электронное обучение и дистанционные образовательные технологии.

6.1. Образовательные технологии

Учебные занятия по дисциплине могут проводиться с применением информационно-телекоммуникационных сетей при опосредованном (на расстоянии) интерактивном взаимодействии обучающихся и преподавателя в режимах on-line 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

В рамках проведения учебного процесса используется следующий формы проведения учебных занятий:

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

использование виртуальной обучающей среды (или системы управления обучением LMS Moodle) <http://moodle.asu.edu.ru> (размещение учебно-методического материала, публикация заданий для предоставления студентами выполненных отчетов по всем видам работ, ознакомление учащихся с оценками и т.д., размещение объявлений, on-lineконсультации, организация и проведение компьютерного тестирования, обсуждение вопросов в форуме и т.д.), как элемента интерактивного взаимодействия участников образовательного процесса (технологии дистанционного обучения);

использование ресурсов ЭБС и сети Internet, как источников информации;

использование электронных учебников и различных сайтов, как источников информации;

использование возможностей электронной почты преподавателя;

При реализации различных видов учебной и внеучебной работы используются и иные информационные системы, сервисы и мессенджеры.

6.3. Перечень программного обеспечения и информационных справочных систем

а) Перечень лицензионного учебного программного обеспечения:

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

б) 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>

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) «Программирование для анализа данных» проверяется сформированность у обучающихся компетенций, указанных в разделе 3 настоящей программы. Этапность формирования данных компетенций в процессе освоения образовательной программы определяется последовательным освоением дисциплин (модулей) и прохождением практик, а в процессе освоения дисциплины (модуля) – последовательным достижением результатов освоения содержательно связанных между собой разделов, тем.

Таблица 5
Соответствие разделов, тем дисциплины (модуля),
результатов обучения по дисциплине (модулю) и оценочных средств

№ п/п	Контролируемые разделы, темы дисциплины (модуля)	Код контролируемой компетенции (компетенций)	Наименование оценочного средства
1	Основы программирования на языке Python (refresher).	ОПК-6, ПК-2	Лабораторная работа №1
2	Математический аппарат (refresher).	ОПК-6, ПК-2	Лабораторная работа №2
3	Введение в модуль NumPy.	ОПК-6, ПК-2	Лабораторная работа №3
4	Основы работы с Pandas.	ОПК-6, ПК-2	Лабораторная работа №4
5	Разведывательный анализ данных.	ОПК-6, ПК-2	Лабораторная работа №5
6	Визуализация данных. Представление результатов	ОПК-6, ПК-2	Лабораторная работа №6
7	Работа с текстовыми данными. Текстовый анализ.	ОПК-6, ПК-2	Лабораторная работа №7. Вопросы к экзамену

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

Таблица 6
Показатели оценивания результатов обучения в виде знаний

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует глубокое знание теоретического материала, умение обоснованно излагать свои мысли по обсуждаемым вопросам, способность

	полно, правильно и аргументированно отвечать на вопросы, приводить примеры
4 «хорошо»	демонстрирует знание теоретического материала, его последовательное изложение, способность приводить примеры, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует неполное, фрагментарное знание теоретического материала, требующее наводящих вопросов преподавателя, допускает существенные ошибки в его изложении, затрудняется в приведении примеров и формулировке выводов
2 «неудовлетворительно»	демонстрирует существенные пробелы в знании теоретического материала, не способен его изложить и ответить на наводящие вопросы преподавателя, не может привести примеры

Таблица 7

Показатели оценивания результатов обучения в виде умений и владений

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы
4 «хорошо»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует отдельные, несистематизированные навыки, не способен применить знание теоретического материала при выполнении заданий, испытывает затруднения и допускает ошибки при выполнении заданий, выполняет задание при подсказке преподавателя, затрудняется в формулировке выводов
2 «неудовлетворительно»	не способен правильно выполнить задание

7.3. Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности

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 «Древо 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`.

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

а) 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>

б) 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>

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

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

При необходимости рабочая программа дисциплины (модуля) может быть адаптирована для обеспечения образовательного процесса инвалидов и лиц с ограниченными возможностями здоровья, в том числе для обучения с применением дистанционных образовательных технологий. Для этого требуется заявление студента (его законного представителя) и заключение психолого-медико-педагогической комиссии (ПМПК).