

ФГБОУ ВО Новосибирский ГАУ
Факультет экономики и управления
Кафедра Информационных технологий и моделирования

ВИЗУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Методические указания для проведения лабораторно-практических занятий и
самостоятельной работы студентов

Новосибирск 2021

УДК 004.42 (07)
ББК 32.972, я 7
В 428

Кафедра Информационных технологий и моделирования

Составитель: И.С. Казакова, старший преподаватель кафедры Информационных технологий и моделирования

Рецензент: Е.Ю. Давыдова, старший преподаватель кафедры Учета и финансовых технологий

Визуальный анализ данных: методические указания для проведения лабораторно-практических занятий и самостоятельной работы студентов/ Новосибир. гос. аграр. ун-т. Фак. ЭиУ; сост. И.С. Казакова - Новосибирск, 2021. – 8 с.

Методические указания для лабораторно-практических занятий и самостоятельной работы студентов по дисциплине «Визуальный анализ данных» предназначены для студентов направления подготовки 09.03.03 Прикладная информатика всех форм обучения.

Методические указания утверждены и рекомендованы к изданию учебно-методическим советом факультета экономики и управления (протокол №4 от «28» декабря 2021 г.)

© Новосибирский государственный аграрный университет, 2021

1. Введение

Лабораторно-практическое занятие – форма организации обучения, которая направлена на формирование практических умений и навыков и является связующим звеном между самостоятельным теоретическим освоением обучающимися учебной дисциплины и применением ее положений на практике.

Лабораторно-практические занятия проводятся в целях: выработки практических умений и приобретения навыков в решении задач, выполнении заданий, производстве расчетов, разработке и оформлении документов, практического овладения иностранными языками и компьютерными технологиями. Главным их содержанием является практическая работа каждого обучающегося.

Самостоятельная работа обучающихся – вид деятельности, при котором в условиях систематического уменьшения прямого контакта с преподавателем обучающимися выполняются учебные задания. К таким заданиям относятся контрольные и курсовые работы, рефераты, эссе, доклады и т.д.

При этом специфика самостоятельной работы обучающихся заключается в том, чтобы обучающиеся самостоятельно получали новые знания.

Цель и задачи дисциплины

Целью изучения курса «Визуальный анализ данных» является: являются овладение обучающимися моделями и методами интеллектуального анализа данных и машинного обучения в задачах поиска информации, обработки и анализа данных, а также приобретение навыков исследователя данных и разработчика математических моделей, методов и алгоритмов анализа данных.

Исходя из цели, в процессе изучения дисциплины решаются следующие задачи:

- внедрение и использование прикладного программного обеспечения, реализующего исследование и визуализацию данных;
- сбор и анализ информации бизнес-анализа для формирования возможных решений;
- использование основных моделей и методов машинного обучения и разработки данных;
- применения методов машинного обучения при решении нестандартных задач, в том числе в новой или незнакомой среде.

Дисциплина «Визуальный анализ данных» в соответствии с требованиями ФГОС ВО направлена на формирование следующей компетенции бакалавра.

Профессиональная компетенция (ПК):

- Способен разрабатывать модели бизнес-процессов заказчика с учетом требований к информационным системам (ПК-2).

2. Цель проведения лабораторно-практических занятий и самостоятельной работы обучающихся

Лабораторно-практические занятия направлены на закрепление и расширение знаний, полученных на лекциях.

Лабораторно-практические занятия по курсу «Визуальный анализ данных» направлены на практическое углубленное изучение учебной дисциплины, привитие обучающимся навыков самостоятельного поиска и анализа учебной информации, формирование и развитие у них научного мышления, умения активно участвовать в творческой дискуссии, делать правильные выводы, аргументировано излагать и отстаивать свое мнение.

Цель самостоятельной работы обучающихся – овладение методами получения новых знаний, приобретение навыков самостоятельного анализа социальных явлений и процессов, усиление научных основ практической деятельности.

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

3. Содержание занятий

Тема 1. Библиотека NumPy. Работа с массивами NumPy.

Массивы. Создание массивов с NumPy. Атрибуты массивов. Индексация и "прихотливая" индексация. Срезы. Изменение формы массива. Слияние и разбиение массивов. Универсальные функции: операции над массивами (арифметические, агрегирование, сравнение). Структурированные массивы и массивы записей. Типы данных в NumPy.

Тема 2. Библиотека Pandas для обработки и анализа данных. Обработка данных в Pandas.

Библиотека Pandas. Объект Series. Объект DataFrame. Чтение и запись данных. Объект Index. Иерархическое индексирование. Работа с данными в Pandas: типы данных, арифметические операторы, обработка данных, агрегирование, группировка, работа со строками, соединение данных.

Тема 3. Визуализация данных. Визуализация с помощью библиотек Matplotlib и Seaborn. Возможности библиотеки Pandas для визуализации.

Библиотека Matplotlib. Визуализация данных с помощью библиотеки Matplotlib. Основные компоненты Matplotlib. Размещение текстовых элементов на рисунке в matplotlib. Библиотека Seaborn. Основные типы диаграмм в Matplotlib. Визуализация данных в Seaborn. Возможности библиотеки Pandas для визуализации.

Тема 4. Элементы статистики. Подготовка и исследование данных.

Основы теории вероятностей. Случайные величины и законы их распределения. Непрерывные случайные величины. Введение в математическую статистику. Описательная статистика. Использование Python и его библиотек в описательной статистике. Практическая статистика и визуализация с Python.

4. Контролирующие материалы для аттестации по дисциплине

Темы контрольной работы

1. Оценка принадлежности объектов классам с помощью модели логистической регрессии.
2. Алгоритм k-ближайших соседей для классификации данных.
3. Сегментация клиентской базы с помощью иерархического кластерного анализа.
4. Классификация изображений с помощью CNN: Три марки молока.
5. Технология обработки, анализа и визуализация текстовых данных.
6. Недообучение и переобучение регрессионных моделей.
7. Анализ многомерных данных с помощью метода главных компонент.
8. Предсказание кредитной платёжеспособности клиентов банка методом k-ближайших соседей.
9. Технологии обработки пропущенных значений в данных.
10. Прогнозирование стоимости недвижимости на основе линейной регрессии.

Вопросы к зачету с оценкой

1. Что такое визуализация данных.
2. Классификация по цели представления данных.
3. Визуализация как этап анализа данных.
4. Характеристики средств визуализации данных.
5. Язык Python и особенности его стиля программирования.
6. Синтаксис и управляющие конструкции языка Python. Переменные, значения и их типы. Типы данных в Python.

7. Встроенные операции и функции. Основные алгоритмические конструкции.
8. Списки, кортежи и словари.
9. Наука о данных и Python. Библиотеки: NumPy, Pandas, Matplotlib.
10. Основы NumPy: многомерные массивы и векторные вычисления. Индексирование и вырезание. Универсальные функции: быстрые поэлементные операции над массивами.
11. Обработка данных с применением массивов. Типы данных в NumPy.
12. Визуализация данных в Python. Обзор библиотеки Matplotlib.
13. Возможности библиотеки Pandas для визуализации. Обзор библиотеки Pandas.
14. Возможности библиотеки Pandas для визуализации. Построение простых графиков, таких как гистограммы, графические графики, точечные графики.
15. Работа с данными в Pandas: типы данных, арифметические операторы, обработка данных, агрегирование, группировка, работа со строками, соединение данных.
16. Визуализация данных в Seaborn.
17. Основы математической статистики. Сбор, обработка и анализ данных с помощью Python.
18. Введение в анализ табличных данных в Python. Пакет pandas. Объекты Series (последовательность) и DataFrame (таблица).

5. Рекомендуемая литература

Основная литература

1. Коган, Е. А.. Теория вероятностей и математическая статистика: учебник / Е. А. Коган, А. А. Юрченко. - Москва: ИНФРА-М, 2020. - 250 с. - (Высшее образование: Бакалавриат). - ISBN 978-5-16-014235-7. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1052969>

2. Григорьев, А. А. Передача, хранение и обработка больших объемов научных данных: учебное пособие / А.А. Григорьев, Е.А. Исаев, П.А. Тарасов. — Москва: ИНФРА-М, 2021. — 207 с. — (Высшее образование: Бакалавриат). — DOI 10.12737/1073525. - ISBN 978-5-16-015985-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1073525>

Дополнительная литература

1. Двойцова, И. Н. Элементы теории вероятностей и математической статистики: учебное пособие / И. Н. Двойцова. - Железногорск: ФГБОУ ВО Сибирская пожарно-спасательная академия ГПС МЧС России, 2021. - 136 с. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1844137>

2. Математическая статистика: практикум / сост. Н. А. Андреева, Р. В. Кузьменко, Е. В. Корчагина, Т. В. Меньших; ФКОУ ВО Воронежский институт ФСИН России. - Воронеж: Научная книга, 2020. - 103 с. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1240998>

Информационное обеспечение

1. Электронно-библиотечная система издательства «Лань» // www.e.lanbook.com.

2. Электронно-библиотечная система издательства «ИНФРА-М» // www.znanium.com.

3. Научная электронная библиотека eLibrary.ru

4. Электронно-библиотечная система НГАУ// <http://nsau.edu.ru/library/ebooks/e-lib-sys-nsau/>

5. Библиотечно-информационный ресурс «Федеральное собрание образовательных материалов для студентов».

Составитель
Казакова Ирина Сергеевна

Методические указания для проведения лабораторно-практических занятий и
самостоятельной работы студентов

Новосибирский государственный аграрный университет

630039, Новосибирск, ул. Добролюбова, 16