Лабораторная работа № 1

ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА МАШИННОГО ОБУЧЕНИЯ

Цель работы: изучение программных средств для организации рабочего места специалиста по анализу данных и машинному обучению

Задача: установка и настройка среды разработки Python; изучение принципов загрузки и очистки данных; получение навыков по предварительной обработке данных на языке Python; изучение основных библиотек Python для работы с данными

**Описание набора данных**

Был скачан набор данных из репозитория Center for Machine Learning and Intelligent Systems: https://archive.ics.uci.edu/dataset/222/bank+marketing

Файл содержит данные о банковских клиентах с различными категориями:

1. Данные банковского клиента:

1 - возраст (числовой)

2 - работа: тип работы (категориальный: "администратор", "неизвестно", "безработный", "руководство", "домохозяйка", "предприниматель", "студент", "синий воротничок", "самозанятый", "пенсионер", "технический специалист", "услуги")

3 - семейное положение: семейное положение (категориальное: "женат", "разведен", "холост"; примечание: "разведен" означает разведен или вдовец)

4 - образование (категориальное: "неизвестно", "среднее", "начальное", "высшее")

5 - кредитный долг: есть ли кредит в кредитный долг? (двоичный: "да", "нет")

6 - баланс: средний годовой баланс в евро (числовой)

7 - жилье: есть ли жилищный кредит? (двоичный: "да", "нет")

8 - кредит: есть ли личный кредит? (двоичный: "да", "нет")

связано с последним контактом текущей кампании:

9 - контакт: тип связи контакта (категориальный: "неизвестно", "телефон", "сотовый")

10 - день: день последнего контакта месяца (числовой)

11 - месяц: месяц последнего контакта года (категориальный: "янв", "фев", "март", ..., "ноя", "дек")

12 - продолжительность: продолжительность последнего контакта в секундах (числовой)

1. Другие атрибуты:

13 - кампания: количество контактов, выполненных во время этой кампании и для этого клиента (числовой, включая последний контакт)

14 - предыдущие дни: количество дней, прошедших после последнего контакта с клиентом из предыдущей кампании (числовой, -1 означает, что с клиентом ранее не связывались)

15 - предыдущие: количество контактов, выполненных до этой кампании и для этого клиента (числовой)

16 - результат предыдущей: результат предыдущей маркетинговой кампании (категориальный: "неизвестно", "другое", "неудача", "успех")

1. Выходная переменная (желаемая цель):

17 - y - подписался ли клиент на срочный депозит? (двоичный: "да", "нет")

**Форма набора данных**

Количество элементов набора: 4521

Количество признаков: 17

Количество пропущенных значений: 4021

Среднее значение отдельных признаков, максимальные и минимальные значения отдельных признаков:

1. Возраст: средний – 41, максимальный – 87, минимальный – 19.
2. Баланс: средний – 1422, максимальный – 71188, минимальный – ‑3133.
3. Продолжительность: средний ­– 263, максимальный – 3025, минимальный – 4

**Предположения исходя из графиков**

На каждом графике красными точками отмечены те клиенты, которые согласились на предложение.

На рисунке 1 показана, что клиент охотна соглашается если он младше 60 лет и его баланс меньше 5000$.

Изображение выглядит как снимок экрана, текст, диаграмма, линия

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 1 – Анализ согласия клиента от баланса и возраста

На рисунке 2, видно, что клиент может согласиться если также его баланс меньше 5000$ и продолжительность разговора ≥ 150 секунд.

Изображение выглядит как снимок экрана, текст, диаграмма, Красочность

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 2 – Анализ согласия клиента от продолжительности и баланса

На рисунке 3, видно, что клиент согласиться, если продолжительность разговора ≥ 150 секунд и его возраст меньше 62 лет.

Изображение выглядит как текст, снимок экрана, диаграмма, линия

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 2 – Анализ согласия клиента от возраста и продолжительности

Приложение 1

|  |
| --- |
| def f(file):  a = []  while True:  content = file.readline()  if not content:  break  try:  a.append(int(content.strip()))  except ValueError:  print(f"Ошибка преобразования строки '{content.strip()}' в число.")  return a |
| import matplotlib.pyplot as plt  def vis(mas1, mas2, color, strr, str1, str2):  plt.figure(figsize=(10, 6))  plt.title(strr)  plt.xlabel(str1)  plt.ylabel(str2)  plt.scatter(mas1, mas2, color=color) |
| file = open('D:/py anal data/bank/y.txt', 'r')  y = f(file)  y = y[:501]  file.close()  file = open('D:/py anal data/bank/возраст.txt', 'r')  v = f(file)  v = v[:501]  file.close()  file = open('D:/py anal data/bank/баланс.txt', 'r')  bal = f(file)  bal = bal[:501]  file.close()  file = open('D:/py anal data/bank/прод.txt', 'r')  pr = f(file)  pr = pr[:501]  file.close()  colors = ['blue'] \* len(bal)  for i, value in enumerate(bal):  if y[i] == 1:  colors[i] = 'red'  vis(v, bal, colors, "Анализ от баланса и возраста", "Возраст", "Баланса")  vis(pr, bal, colors, "Анализ от продолжительности и баланса", "Продолжительность (сек)", "Баланс")  vis(v, pr, colors, "Анализ от возраста и продолжительности", "Возраст", "Продолжительность (сек)")  plt.show() |