МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

ФГАОУ ВО «СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

ИНСТИТУТ ИНФОРМАЦИОННЫХ

ТЕХНОЛОГИЙ И ТЕЛЕКОММУНИКАЦИЙ

КАФЕДРА ПРИКЛАДНОЙ ИНФОРМАТИКИ

**Лабораторная работа №7**

по дисциплине

«Искусственный интеллект и машинное обучение»

**Выполнил:**

Сердюков Никита Анатольевич

Студент 2 курса группы \_ПИН-б-о-22-1

Направления подготовки

09.03.03 Прикладная информатика

очной формы обучения

Ставрополь, 2023 г.

Тема: ЛОГИЧЕСКИЕ МЕТОДЫ КЛАССИФИКАЦИИ.

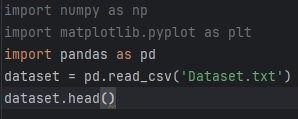
Цель работы: изучение теоретических принципов и инструментальных средств для построения пайплайна для предварительной обработки данных.

Выполнение работы:

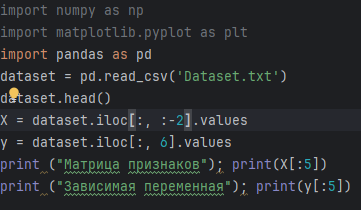
Датасет: Smartphone classification

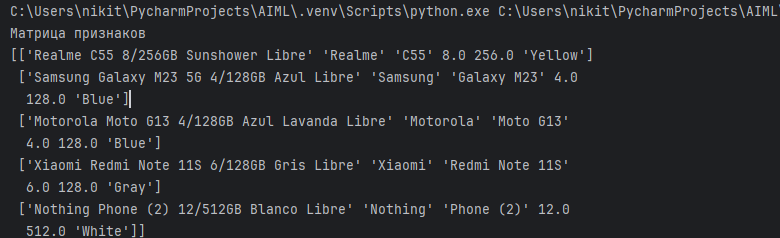
Описание датасета: Датасет содержит информацию о различных смартфонах, позволяя провести детальный анализ их технических характеристик и цен. Он охватывает широкий спектр смартфонов различных брендов, моделей и конфигураций. Признаки: Смартфон, Бренд, Модель, ОЗУ, Хранилище, Бесплатная сим-карта(по контракту), Финальная цена

1. Загрузка данных

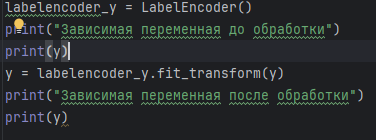


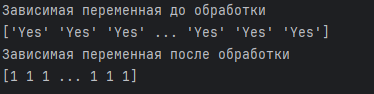
1. Разделение на матрицу признаков и зависимую переменную

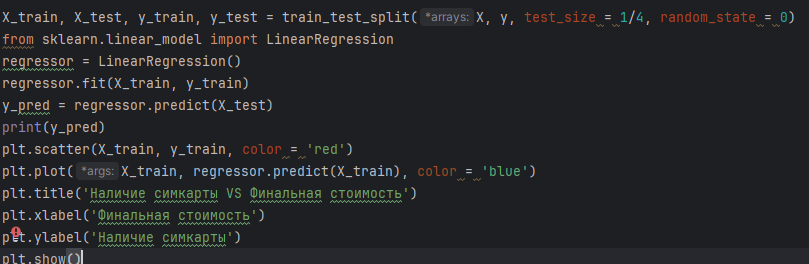


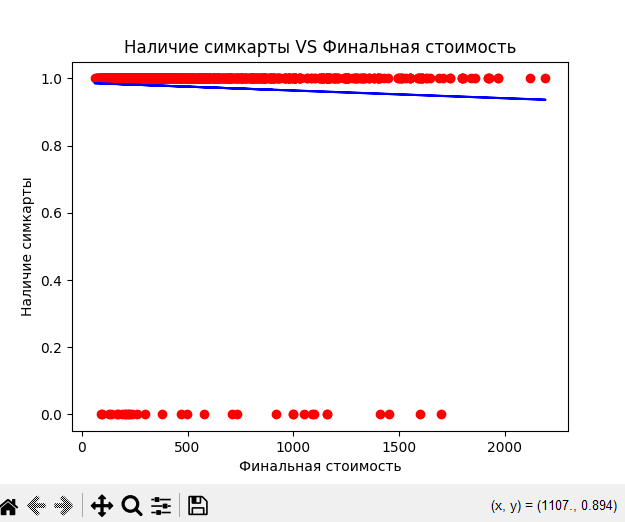


1. Обработка категориальных данных





1. Разделение выборки на тестовую и тренировочную, визуализация.



Листинг:

import numpy as np  
import matplotlib.pyplot as plt  
from sklearn.model\_selection import train\_test\_split  
import pandas as pd  
from sklearn.preprocessing import LabelEncoder  
dataset = pd.read\_csv('Dataset.txt')  
dataset.head()  
X = dataset.iloc[:,-1:].values  
y = dataset.iloc[:, 6].values  
print ("Матрица признаков"); print(X[:5])  
print ("Зависимая переменная"); print(y[:5])  
  
  
labelencoder\_y = LabelEncoder()  
print("Зависимая переменная до обработки")  
print(y)  
y = labelencoder\_y.fit\_transform(y)  
print("Зависимая переменная после обработки")  
print(y)  
  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 1/4, random\_state = 0)  
from sklearn.linear\_model import LinearRegression  
regressor = LinearRegression()  
regressor.fit(X\_train, y\_train)  
y\_pred = regressor.predict(X\_test)  
print(y\_pred)  
plt.scatter(X\_train, y\_train, color = 'red')  
plt.plot(X\_train, regressor.predict(X\_train), color = 'blue')  
plt.title('Наличие симкарты VS Финальная стоимость')  
plt.xlabel('Финальная стоимость')  
plt.ylabel('Наличие симкарты')  
plt.show()

Контрольные вопросы:

1. Библиотека pandas

2. б) удаление строк, содержащих пропуски в данных

3. Нет, так как категориальные переменные уже представлены в виде отдельных классов.

4. Наиболее оптимальное соотношение тестовой и обучающей выборки - 80:20.

5. а) dataset = read\_csv("data.csv")

Вывод: изучил принципы построения информационных систем с использованием логических методов классификации.