**Digital Assessment 4**

**CBS3007 -** **Data Mining and Analytics**

Date: 3 November, 2024

Name: Anuj Parihar

Registration Number: 21BBS0162

**Link to Assessment Codebase and Dataset**:

**Question 1**

Consider a dataset of 50 user records with the attributes “Name”, “location” ,“Height”,

“Weight”, ”Age”. Do the following tasks.

i) Create the dataset for the attributes given.

ii) Implement the Demo on Classification Technique using KNN.

**Aim:** The aim of this project is to implement a K-Nearest Neighbors (KNN) classification technique on a synthetic dataset of user information. The dataset includes the attributes Name, Location, Height, Weight, and Age for 50 individuals.

**Sample Input:** The entire input dataset is in the GitHub repository

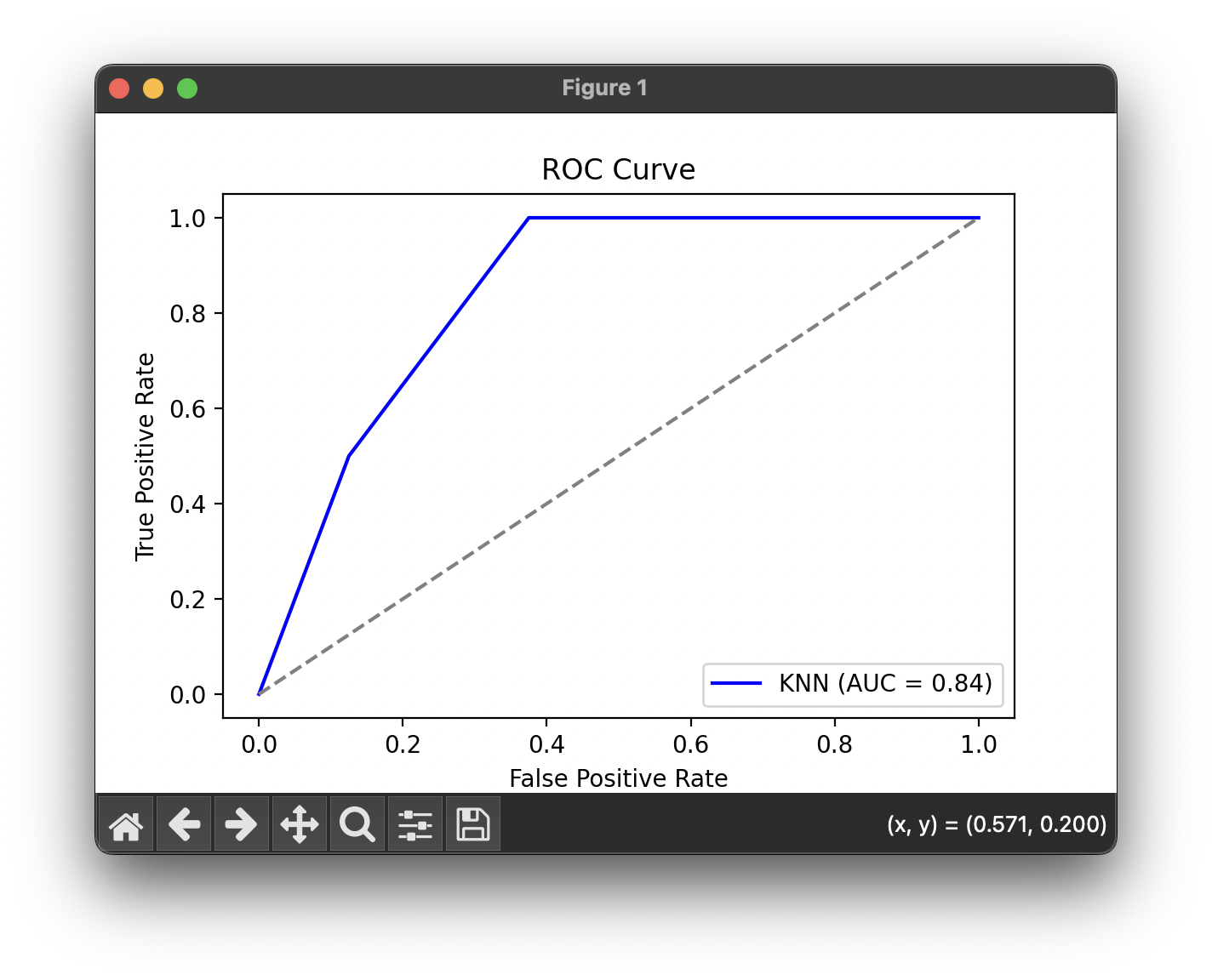
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Location | Height | Weight | Age |
| Julie King | Bengaluru | 159 | 55 | 40 |
| Brenda Martinez | Delhi | 172 | 92 | 20 |
| Carlos Miller | Indore | 167 | 88 | 32 |
| Mary Barnett | Delhi | 164 | 74 | 45 |
| Caitlin Doyle | Indore | 168 | 51 | 25 |
| Chad Wolfe | Mumbai | 182 | 82 | 28 |
| Brian Herring | Delhi | 174 | 63 | 22 |
| Jennifer Smith | Vellore | 170 | 59 | 35 |
| Edward Lane | Delhi | 165 | 68 | 24 |
| Elizabeth Ramirez | Delhi | 172 | 73 | 18 |

**Output:**

**A computer screen with white text

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**Results:**

The KNN classifier successfully classified individuals as "Overweight" or "Not Overweight" based on BMI with an accuracy of 0.80. Key metrics, including precision, recall, F1 score, and ROC AUC, indicate reliable performance, demonstrating the model's effectiveness in distinguishing between the two classes.

**Question 2:**

Linear regression of 2 variables is to use one variable to forecast another variable value.

**Sample Input:** The entire input is in the GitHub Repository

**Output:**

**Results:**