InClass Assignment 5

Rosemarie Kpaka

Bowie State University

CTEC402-101-Fall2024 Software and OS Security

Professor Adrian D. Carter

November 3, 2024

**Part 1: Machine Learning Algorithms Explanation**

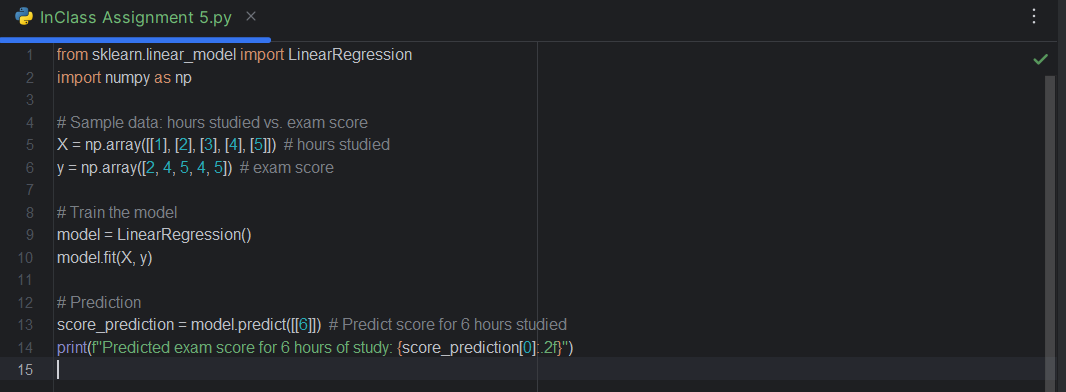
Below, I will explain each selected Machine Learning algorithm in my own words. I’ll also include ChatGPT’s explanations side-by-side for comparison.

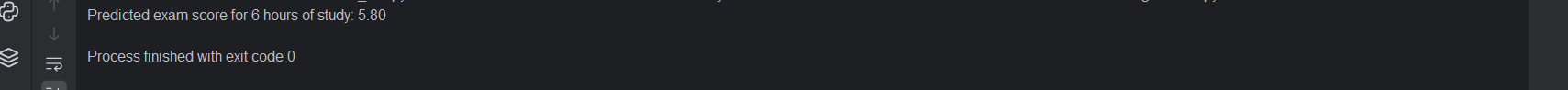
|  |  |  |
| --- | --- | --- |
| **Algorithm** | **My Explanation** | **ChatGPT Explanation** |
| 1. Linear Regression | Linear regression is a simple algorithm that is used to predict a continuous outcome. It aims to draw a line through points that minimizes the distance-error between the line and the actual points. It works by drawing the best-fit line, represented by the equation y = mx + b, where m is the slope and b is the y-intercept. | Linear regression is a statistical method that models the relationship between a dependent variable and one or more independent variables by fitting a straight line to the data. This line minimizes the error between predicted and actual values. |
| 2. Logistic Regression | Logistic regression is the type of regression used when there is a binary classification problem. Other than fitting a line, it uses a logistic function that outputs a value between 0 and 1, which can then be used to classify data points as one class or another based on some threshold, say 0.5. | Logistic regression is a classification algorithm that predicts probabilities for binary outcomes. It uses a logistic function to model the data, providing an output between 0 and 1 to decide the class label. |
| 3. Decision Tree | A Decision tree assigns a tree-like model of decisions. Further, it splits data into branches by means of yes/no questions based on features values. At each "node", the algorithm picks the best feature to split the data further down until the decision point is reached at the "leaf" node. | Decision trees split data into branches based on features, working through a series of questions to reach a decision at leaf nodes. It repeatedly splits the data, aiming to minimize classification error at each step. |
| 4. SVM (Support Vector Machine) | The algorithm of SVM classification finds the hyperplane such that classes are best separated. It tries to maximize the margin or distance between various classes in data and effectively creates a "decision boundary" which divides classes with the largest gap possible. | Support Vector Machines separate data into classes by finding a hyperplane that maximizes the margin between classes. It selects support vectors to ensure the largest possible gap between classes for accurate separation. |
| 5. K-means Clustering | K-means is an unsupervised algorithm that segregates a dataset into groups, basically based on similarities. It starts with choosing a random number k of centroids, henceforth mapping each point to the closest centroid. Thereafter, it enters into a recursive process of updating the centroid and the cluster so that all points are as near as possible to the centroid of their cluster. | K-means clustering assigns data into k clusters based on proximity to centroids. The centroids are iteratively adjusted to minimize the distance between points within each cluster. |

**Part 2: Programming Two Machine Learning Algorithms**

Below are Python implementations of Linear Regression and Decision Tree algorithms. Both scripts will train models on sample data and display predictions.

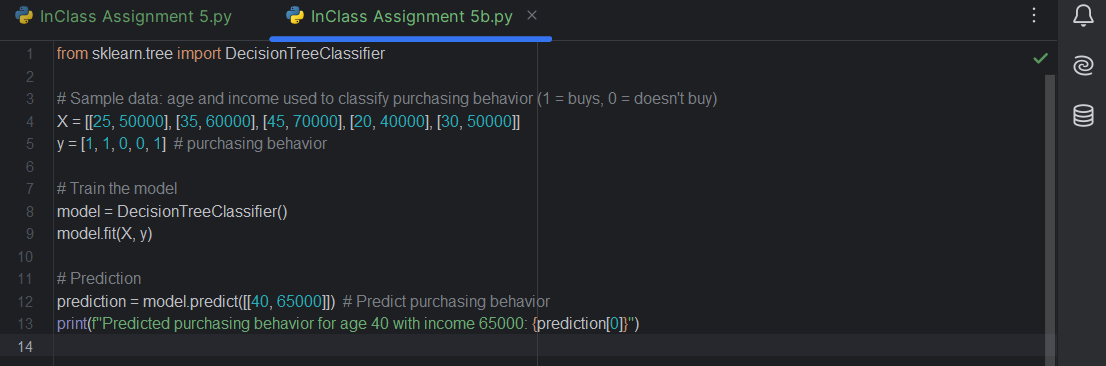
1. Linear Regression (Python)

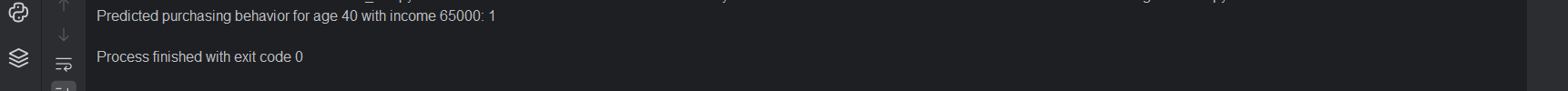




The model is trained on sample data, and then it predicts the exam score for a student who studied for 6 hours.

1. Decision Tree (Python)





The Decision Tree model predicts whether a person of age 40 with a $65,000 income would make a purchase.