

Student Performance Predictor

Shreyas Ghuge
sghug001@ucr.edu

Naman Mittal
nmitt005@ucr.edu

Imran Shahid
ishah011@ucr.edu

Madhusudhan Tungamitta
mtung007@ucr.edu

Balaji Arunachalam
barun001@ucr.edu

Introduction

- In most educational entities and institutes, predicting student performance has become a pressing need.
- This is necessary in order to assist at-risk students and ensure their retention, and provide exceptional learning resources and experiences
- Attributes such as school, age, sex, attendance, health, previous grade record etc are reasonable attributes to understand and predict a student's upcoming grades.
- We are trying to devise a system that will help students recognize their final grades and improve their academic behavior.
- Furthermore, using and writing multiple machine learning models we try to compare the performance and output based on a key measure like accuracy.

Proposed Method

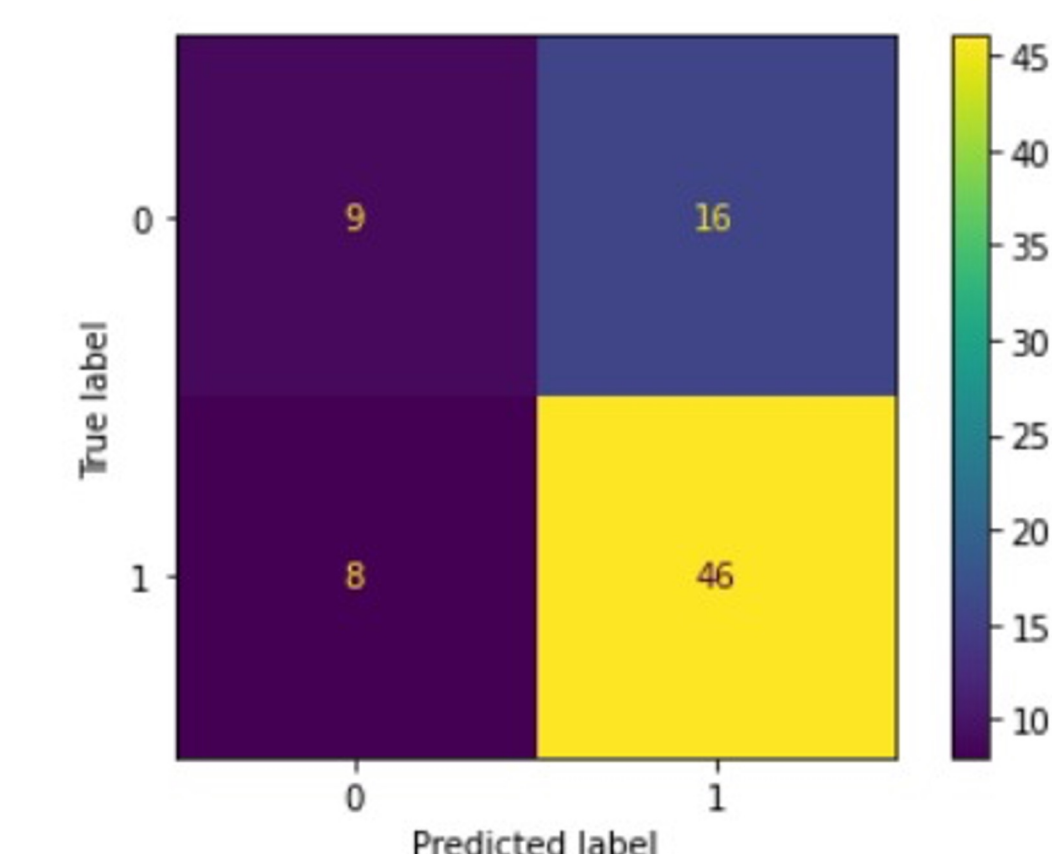
- We started with preprocessing the input data. Preprocessing steps include removing null values, encoding values to numerical format and applying scaling for efficient calculations.
- We tried and implemented 5 Machine Learning algorithms to imitate the working of the one's that Scikit Learn provides along with the goal of successfully predicting if a student will pass or fail given certain academic parameters.
- We then implemented each of the 5 algorithms from scratch using python and compared accuracies from scikit learn. We were able to achieve accuracy as high as 68% using Naive Bayes Algorithm.

Problem Definition

- We compare the fulfillment of several Machine algorithms to the one's provided by python's inbuilt Scikit Learn library.
- The fundamental problem definition revolves around comparing the ML models created using Scikit Learn versus the one's implemented from scratch.
- Supervised learning algorithms implemented were Random Forest, Decision Trees, Support Vector Machines, naive bayes and Logistic regression. We could match the accuracy of the above algorithms upto 95%.

Results

- We used 4 metrics namely accuracy, precision, recall, and F1 score for comparing the classifiers.
- All of the machine learning models had comparable findings, as we expected. On the test sample accuracies ranged from 0.6 to 0.7.
- Decision Tree Algorithm - Accuracy: 62 %
- Random Forest Algorithm - Accuracy: 63.29 %
- SVM Algorithm - Accuracy: 60.76 %
- Naive Bayes Algorithm - Accuracy: 68.35 %
- Logistic Regression - Accuracy: 64.5 %



Related Work

There have been multiple instances of thorough research conducted on this topic. Many of them used different methods to solve the titled problem statement and hence computed relevant scores to rate the working.

1. Authors J. Dhilipan et. al
2. Sitha Ram and team
3. Yağcı, M. Educational data mining: prediction of students' academic
4. Functional near-infrared spectroscopy to determine students' performance.

