MINI PROJECT – II
(2019-20)

Student Performance Prediction

MID TERM REPORT

Institute of Engineering & Technology

Team Members

Akshay Dubey (171500029)
Himanshu Pathak (171500136)
Ganesh Pandey (171500106)
Saurabh Singhal (171500304)

Supervised By
Mr. Vinay Agrawal
(Assistant Professor)

Department of Computer Engineering & Applications

ACKNOWLEDGEMENT

We have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. We would like to extend our sincere thanks to all of them.

We are highly indebted to **Mr. Vinay Agrawal** for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

We would like to express our special gratitude and thanks to industry persons for giving me such attention and time.

Our thanks and appreciations also go to my colleague in developing the project and people who have willingly helped us out with their abilities.

CERTIFICATE

This is to certify that Akshay Dubey, Himanshu Pathak, Ganesh Pandey, Saurabh Singhal students of B.Tech(CSE) 3rd Year , GLA University , Mathura have successfully completed the Project on "Student Performance Prediction" under the guidance of "Mr. Vinay Agrawal", Assistant Professor GLA University, Mathura.

Mr. Vinay Agrawal
(Signature)

Problem Statement

Universities gather large volumes of data with reference to their students in electronic form. The advances in the data mining field make it possible to mine these educational data and find information that allow for innovative ways of supporting both teachers and students. The results show that it is possible to predict the graduation performance in 4th year at university using only pre-university marks and marks of 1st and 2nd year courses, no socio-economic or demographic features, with a reasonable accuracy. Furthermore courses that are indicators of particularly good or poor performance have been identified.

Reason for selecting the topic

To help Teachers and Students both in which area they are lagging. From the point of teachers it will help them by showing them the students to whom they should pay more attention. For students it will show them the subjects in which they are lagging or they should apply more efforts in that field.

Objective of the Project

The following tasks that will be performed by this tool are :
* Teacher/Admin can add the data of students.
* Student/Teacher both can see Students report.
* Analyze the Student's report.
* Shows teacher the students they will pay more attention.
* Show Student's predicted marks.

Future Scope

It will act as an aid for the universities or institutions as universities gather large volumes of data with reference to their students in electronic form. The advances in the data mining field make it possible to mine these educational data and find information that allow for innovative ways of supporting both teachers and students. It will help students to improve their overall grade and also the university or institutions to improve their overall result.

Hardware & Software to be used

Hardware:
* Laptop with 8GB RAM

Software:
* Python
* ML

What contribution would the project make and where

This project will make contribution in universities and institutions with great accuracy to improve their students' performance. It will also help students in making their future bright. The massive growth in the educational sector needs to create awareness about handling the huge volume of student data. The educational data mining is a technique to extract information from these volumes of data. Nowadays educational data mining technique plays a vital role in predicting academic performance.

Scope for extension into a major project

As the tech is increasing day by day, so use of this technology we can provide the instant analysis of their performance to all the students. This project can be extended as proper analysis platform for students and suggesting them the video lectures for the subjects they have to focus. So, this project can contribute in making our education pattern better.

Dataset

```
school;sex;age;address;famsize;Pstatus;Medu;Fedu;Mjob;Fjob;reason;guardian;traveltime;studytime;failures;schoolsup;famsup;paid;activities;nursery;higher;internet;romantic;famrel;freetime;goout;Dalc;Walc;health;absences;G1;G2;G3
GP;"F";18;"U";"GT3";"A";4;4;"at_home";"teacher";"course";"mother";2;2;0;"yes";"no";"no";"no";"yes";"yes";"no";"no";4;3;4;1;1;3;6;"5";"6";6
GP;"F";17;"U";"GT3";"T";1;1;"at_home";"other";"course";"father";1;2;0;"no";"yes";"no";"no";"no";"yes";"yes";"no";5;3;3;1;1;3;4;"5";"5";6
GP;"F";15;"U";"LE3";"T";1;1;"at_home";"other";"other";"mother";1;2;3;"yes";"no";"yes";"no";"yes";"yes";"yes";"no";4;3;2;2;3;3;10;"7";"8";10
GP;"F";15;"U";"GT3";"T";4;2;"health";"services";"home";"mother";1;3;0;"no";"yes";"yes";"yes";"yes";"yes";"yes";"yes";3;2;2;1;1;5;2;"15";"14";15
GP;"F";16;"U";"GT3";"T";3;3;"other";"other";"home";"father";1;2;0;"no";"yes";"yes";"no";"yes";"yes";"no";"no";4;3;2;1;2;5;4;"6";"10";10
GP;"M";16;"U";"LE3";"T";4;3;"services";"other";"reputation";"mother";1;2;0;"no";"yes";"yes";"yes";"yes";"yes";"yes";"no";5;4;2;1;2;5;10;"15";"15";15
GP;"M";16;"U";"LE3";"T";2;2;"other";"other";"home";"mother";1;2;0;"no";"no";"no";"no";"yes";"yes";"yes";"no";4;4;4;1;1;3;0;"12";"12";11
GP;"F";17;"U";"GT3";"A";4;4;"other";"teacher";"home";"mother";2;2;0;"yes";"yes";"no";"no";"yes";"yes";"no";"no";4;1;4;1;1;1;6;"6";"5";6
GP;"M";15;"U";"LE3";"A";3;2;"services";"other";"home";"mother";1;2;0;"no";"yes";"yes";"no";"yes";"yes";"yes";"no";4;2;2;1;1;1;0;"16";"18";19
GP;"M";15;"U";"GT3";"T";3;4;"other";"other";"home";"mother";1;2;0;"no";"yes";"yes";"yes";"yes";"yes";"yes";"no";5;5;1;1;1;5;0;"14";"15";15
GP;"F";15;"U";"GT3";"T";4;4;"teacher";"health";"reputation";"mother";1;2;0;"no";"yes";"yes";"no";"yes";"yes";"yes";"no";3;3;3;1;2;2;0;"10";"8";9
GP;"F";15;"U";"GT3";"T";2;1;"services";"other";"reputation";"father";3;3;0;"no";"yes";"no";"no";"yes";"yes";"yes";"no";5;2;2;1;1;4;4;"10";"12";12
GP;"M";15;"U";"LE3";"T";4;4;"health";"services";"course";"father";1;1;0;"no";"yes";"yes";"yes";"yes";"yes";"yes";"no";4;3;3;1;3;5;2;"14";"14";14
GP;"M";15;"U";"GT3";"T";4;3;"teacher";"other";"course";"mother";2;2;0;"no";"yes";"yes";"no";"yes";"yes";"yes";"no";5;4;3;1;2;3;2;"10";"10";11
GP;"M";15;"U";"GT3";"A";2;2;"other";"other";"home";"other";1;3;0;"no";"yes";"no";"no";"yes";"yes";"yes";"yes";4;5;2;1;1;3;0;"14";"16";16
GP;"F";16;"U";"GT3";"T";4;4;"health";"other";"home";"mother";1;1;0;"no";"yes";"no";"no";"yes";"yes";"yes";"no";4;4;4;1;2;2;4;"14";"14";14
GP;"F";16;"U";"GT3";"T";4;4;"services";"services";"reputation";"mother";1;3;0;"no";"yes";"yes";"yes";"yes";"yes";"yes";"no";3;2;3;1;2;2;6;"13";"14";14
GP;"F";16;"U";"GT3";"T";3;3;"other";"other";"reputation";"mother";3;2;0;"yes";"yes";"no";"no";"yes";"yes";"no";"no";5;3;2;1;1;4;4;"8";"10";10
GP;"M";17;"U";"GT3";"T";3;2;"services";"services";"course";"mother";1;1;3;"no";"yes";"no";"yes";"yes";"yes";"yes";"no";5;5;5;2;4;5;16;"6";"5";5
GP;"M";16;"U";"LE3";"T";4;3;"health";"other";"home";"father";1;1;0;"no";"no";"yes";"yes";"yes";"yes";"yes";"no";3;1;3;1;3;5;4;"8";"10";10
GP;"M";15;"U";"GT3";"T";4;3;"teacher";"other";"reputation";"mother";1;2;0;"no";"no";"no";"no";"yes";"yes";"yes";"no";4;4;1;1;1;1;0;"13";"14";15
GP;"M";15;"U";"GT3";"T";4;4;"health";"health";"other";"father";1;1;0;"no";"yes";"yes";"no";"yes";"yes";"yes";"no";5;4;2;1;1;5;0;"12";"15";15
GP;"M";16;"U";"LE3";"T";4;2;"teacher";"other";"course";"mother";1;2;0;"no";"no";"no";"yes";"yes";"yes";"yes";"no";4;5;1;1;3;5;2;"15";"15";16
GP;"M";16;"U";"LE3";"T";2;2;"other";"other";"reputation";"mother";2;2;0;"no";"yes";"no";"yes";"yes";"yes";"yes";"no";5;4;4;2;4;5;0;"13";"13";12
GP;"F";15;"R";"GT3";"T";2;4;"services";"health";"course";"mother";1;3;0;"yes";"yes";"yes";"yes";"yes";"yes";"yes";"no";4;3;2;1;1;5;2;"10";"9";8
GP;"F";16;"U";"GT3";"T";2;2;"services";"services";"home";"mother";1;1;2;"no";"yes";"yes";"no";"no";"yes";"yes";"no";1;2;2;1;3;5;14;"6";"9";8
GP;"M";15;"U";"GT3";"T";2;2;"other";"other";"home";"mother";1;1;0;"no";"yes";"yes";"no";"yes";"yes";"yes";"no";4;2;2;1;2;5;2;"12";"12";11
GP;"M";15;"U";"GT3";"T";4;2;"health";"services";"other";"mother";1;1;0;"no";"no";"yes";"no";"yes";"yes";"yes";"no";2;2;4;2;4;1;4;"15";"16";15
GP;"M";16;"U";"LE3";"A";3;4;"services";"other";"home";"mother";1;2;0;"yes";"yes";"no";"yes";"yes";"yes";"yes";"no";5;3;3;1;1;5;4;"11";"11";11
```

Code

```python
import numpy as np
import pandas as pd

df = pd.read_csv(r"C:\Users\User\.spyder-py3\Student Performance Prediction\student-mat.csv", sep=";")

from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import cross_val_score

from sklearn.pipeline import Pipeline
from sklearn.feature_selection import SelectKBest, chi2
from sklearn.svm import LinearSVC

def split_data(X, Y):
    return train_test_split(X, Y, test_size=0.2, random_state=0)

def confuse(y_true, y_pred):
    cm = confusion_matrix(y_true=y_true, y_pred=y_pred)
    print("\nConfusion Matrix: \n", cm)
    fpr(cm)
    ffr(cm)

def fpr(confusion_matrix):
    fp = confusion_matrix[0][1]
    tf = confusion_matrix[0][0]
    rate = float(fp) / (fp + tf)
    print("False Pass Rate: ", rate)

def ffr(confusion_matrix):
    ff = confusion_matrix[1][0]
    tp = confusion_matrix[1][1]
    rate = float(ff) / (ff + tp)
    print("False Fail Rate: ", rate)

    return rate

def train_and_score(X, y):
    X_train, X_test, y_train, y_test = split_data(X, y)

    clf = Pipeline([
        ('reduce_dim', SelectKBest(chi2, k=2)),
        ('train', LinearSVC(C=100))
    ])

    scores = cross_val_score(clf, X_train, y_train, cv=5, n_jobs=2)
    print("Mean Model Accuracy:", np.array(scores).mean())

    clf.fit(X_train, y_train)
    y_pred=clf.predict(X_test)
    confuse(y_test,y_pred)
    print()

def main():
    print("\nStudent Performance Prediction")

    ds=df.drop(columns=['famsize','Pstatus','Mjob','Fjob','reason','guardian','schoolsup','famsup','activities

    class_le = LabelEncoder()
    for column in ds[["school", "sex", "address", "paid", "internet"]].columns:
        ds[column] = class_le.fit_transform(df[column].values)

    for i, row in df.iterrows():
        if row["G1"] >= 10:
            ds["G1"][i] = 1
        else:
            ds["G1"][i] = 0

        if row["G2"] >= 10:
            ds["G2"][i] = 1
        else:
            ds["G2"][i] = 0
```

```python
        if row["G3"] >= 10:
            ds["G3"][i] = 1
        else:
            ds["G3"][i] = 0

    y = ds.pop("G3")

    X = ds

    print("\n\nModel Accuracy Knowing G1 & G2 Scores")
    print("======================================")
    train_and_score(X, y)

    X.drop(["G2"], axis = 1, inplace=True)
    print("\n\nModel Accuracy Knowing Only G1 Score")
    print("======================================")
    train_and_score(X, y)

    X.drop(["G1"], axis=1, inplace=True)
    print("\n\nModel Accuracy Without Knowing Scores")
    print("======================================")
    train_and_score(X, y)
main()
```

Accuracy Scores for different Cases

```
Model Accuracy Knowing G1 & G2 Scores
======================================
Mean Model Accuracy: 0.9111607142857142
C:\Users\User\AppData\Roaming\Python\Pytho
to converge, increase the number of iterat
  "the number of iterations.", Convergence

Confusion Matrix:
 [[28  2]
 [ 4 45]]
False Pass Rate:  0.06666666666666667
False Fail Rate:  0.08163265306122448


Model Accuracy Knowing Only G1 Score
======================================
Mean Model Accuracy: 0.8228174603174601
C:\Users\User\AppData\Roaming\Python\Pytl
to converge, increase the number of iter.
  "the number of iterations.", Convergen

Confusion Matrix:
 [[24  6]
 [ 4 45]]
False Pass Rate:  0.2
False Fail Rate:  0.08163265306122448
```

```
Model Accuracy Without Knowing Scores
=====================================
Mean Model Accuracy: 0.7180555555555556
C:\Users\User\AppData\Roaming\Python\Pyt
to converge, increase the number of iter
  "the number of iterations.", Convergen

Confusion Matrix:
 [[ 8 22]
 [ 2 47]]
False Pass Rate:  0.7333333333333333
False Fail Rate:  0.040816326530612224
```

Conclusion

Universities gather large volumes of data with reference to their students in electronic form. The advances in the data mining field make it possible to mine these educational data and find information that allow for innovative ways of supporting both teachers and students.

It will act as an aid for the universities or institutions as universities gather large volumes of data with reference to their students in electronic form. The advances in the data mining field make it possible to mine these educational data and find information that allow for innovative ways of supporting both teachers and students.

This project can be extended as proper analysis platform for students and suggesting them the video lectures for the subjects they have to focus. So, this project can contribute in making our education pattern better.

References: -

https://archive.ics.uci.edu/ml/datasets/Student+Performance

https://www.researchgate.net/publication/331478516