

# **Student Performance Analysis and Prediction**

*(for the partial fulfillment of Bachelor of Technology Degree in Computer Science & Engineering)*

*Submitted by*  
**Mayank Tyagi**  
**Siddhant Mamgain**  
**Vipin Chand Ramola**  
**Arun Rawat**

Under the guidance of



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**  
**GRAPHIC ERA HILL UNIVERSITY, DEHRADUN**

# CERTIFICATE

This is to certify that the project titled “**Student Performance Analysis and Prediction**” submitted by Mayank Tyagi, Siddhant Mamgain, Vipin Chand Ramola and Arun Rawat of Graphic Era Hill University for the award of the degree of **Bachelor of Technology**, is a bona fide record of the project work done by them under our supervision. The contents of this project in full or in parts have not been submitted to any other Institute or University for the award of any degree or diploma.

**Guide Name**

Project Guide

(Asst. Prof)

GEHU,Dehradun

Place: Dehradun

Date: .....

# ACKNOWLEDGEMENT

It is our privilege to express our sincerest regards to our project coordinator, Guide Name, for his valuable inputs, able guidance, encouragement, whole-hearted cooperation and constructive criticism throughout the duration of our project.

We deeply express our sincere thanks to our Head of Department Dr Navin Garg for encouraging and allowing us to present the project on the topic “Handwritten Character Recognition” at our department premises for the partial fulfillment of the requirements leading to the award of B.Tech degree.

We take this opportunity to thank all our lecturers who have directly or indirectly helped our project. We pay our respects and love to our parents and all other family members for their love and encouragement throughout our career. Last but not the least we express our thanks to our friends for their cooperation and support.

**Mayank Tyagi**

Roll no. 1918468

**Vipin Chand Ramola**

Roll no.1918804

**Siddhant Mamgain**

Roll no. 1918723

**Arun Rawat**

Roll no. 1918262

# Abstract

The academic output of the student is normally stored in various formats in the student administration system (files, documents, records, photographs and other formats). These data can be collected for valuable knowledge from the students. However, it is difficult to analyze the increasing amount of data of students through conventional statistical techniques and database management tools. For universities to gather valuable information, a tool is therefore required. This helpful knowledge can be used to predict the success of students. Leistungs analyze learning results is a framework that aims for success in the areas of student interest at various levels and dimensions. The ultimate goal of any educational institution is offering the best educational experience and knowledge to the students. Identifying the students who need extra support and taking the appropriate actions to enhance their performance plays an important role in achieving that goal. In this research, four machine learning techniques have been used to build a classifier that can predict the performance of the students in a computer science subject. The machine learning techniques include Artificial Neural Network, Naïve Bayes, Decision Tree, and Logistic Regression. This research pays extra attention to the effect of using the internet as a learning resource and the effect of the time spent by students on social networks on the students' performance. These effects introduced by using features that measure whether the student uses the internet for learning and the time spent on the social networks by the students. The models have been compared using the ROC index performance measure and the classification accuracy. In addition, different measures have been computed such as the classification error, precision, recall, and the F measure. The dataset used to build the models is collected based on a survey given to the students and the students' grade book. To seek maximum accuracy in academic predictions across a range of powerful techniques of data mining. The ANN (fully connected feed forward multilayer ANN) model achieved the best performance that is equal to 0.807 and achieved the best classification accuracy that is equal to 77.04%. In addition, the decision tree model identified five factors as important factors which influence the performance of the students.

**Keywords**—Recommender System, Performance Analysis, Statistical Techniques, Decision Tree

# Introduction

Student Performance Analysis's System is an emerging field and is very crucial to schools and universities in helping their student and faculty. As we all know in today's environment, there is a lack of quality education, and also the competition is increasing day by day. so there is need for quality steps to improve the standard of the students. The main goal of every university/School is to provide quality knowledge and skills to students so that they are competitive in the labor market. One way to achieve this goal is to timely predict student performance. The analysis of student's progress during their studies provides the management with the information about the probability of success of each student. Traditionally it is done by the lecturers who use their interaction with students in the class room activities and in the mid-term assessments to identify those at-risk of dropping out and take timely actions. But in the modern system of education, the time of communication between lecturers and student is decreasing and hence it becomes difficult to find weak student or students who need extra attention or support. As in the age of mass computerization, a lot of data is collected in educational institutions, but it remains unused so we can use that data to increase the performance of our students and for this we will be using Machine learning Algorithm (MLA) so that we can take benefit or advantage of the available data.

Many researches have been done on this topic, so researchers had come to a conclusion that technology can be an important factor for analyzing the flaws. And also, the use of technology makes the decision-making easy, as it generates reports and graphs for analysis purpose. So basically in this project we will analyze the performance of the student and then prediction will be done on the data that we got after analysis. We will be able to analyze the performance of single student as well as performance of whole class.

# LITERATURE SURVEY

Many researchers have used several techniques and algorithms in educational institutions for prediction of students' knowledge, particularly decision trees and regression, and finally a lot of ideas have been proposed.

A study was conducted by G.N. Pandey, Sonali Agarwal and M. D. Tiwari in which they added various classifiers to the dataset and tried to find the best classifier by means of comparative analysis.

In the work of Mounika Goyal and Rajan Vokra, techniques like cluster analysis and classification were implemented. Usefulness of OLAP and OLTP were also examined in society. OLTP (Online Transaction Processing) process data in real-time while OLAP (Online Analytical Processing) analyses historical data from OLTP systems.

Further in future many such studies were conducted and ideas were proposed. Different methods can be used to carry on the process.

# TECHNOLOGY STACK

Python, Anaconda and Jupiter.

## 1. PYTHON

Along with being a high-level programming language, python is easy to learn and use as compared to other programming languages. Also, it's developer friendly and has easy to learn basics.

## 2. ANACONDA

Its distribution has already over 250 automatically installed packages. Moreover, 7500 additional packages can also be installed.

## 3. JUPYTER

It allows us to create and share documents. The documents may contain equations, live text, narrative code, etc. It generally has the following uses:

- Numerical simulation
- Data visualization
- Data cleaning and transformation
- Statistical modelling
- Machine learning

# Library Used

- 1) Pandas: Pandas is the software library written for the python programming language it is used for working with datasets. It has functions for analyzing, cleaning, exploring and manipulating data. The name “Pandas” has a referenced to both “Panel data” and “Python Data Analysis” and was created by Wes McKinney in 2008.
- 2) Sklearn : Skit-learn is the most useful and robust library for machine learning in Python.it provides a selection of efficient tools for machine learning and statistical modeling including classification ,regression ,clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy and Matplotlib.
- 3) Matplotlib: it is an amazing visualization library in python for 2D plot of arrays. Matplotlib is a mutli-platform data visualization library built on NumPy .it was introduced by the John Hunter in the year 2002.one of the advantage of this library is that it allows us to access huge amounts of data in visuals.MatPlotlib consists of plots like lines, bar, scatter, histogram etc.
- 4) NumPy: NumPy stands for Numerical Python, is a library consisting of multi-dimensional array objects and a collection of routines for processing those array. Using NumPy, mathematical and logical operations can be performed.



# Problem and Data Understanding

It is one of the most critical steps of building the system and also achieving the objectives and goals. After the problem is well understood, the solution to resolve the problem is found out by studying more relevant research papers.

TABLE I. ATTRIBUTES OF DATASET

Attributes	Values
<b>Gender</b>	Categorical (0=Female, 1=Male)
<b>S1 CGPA</b>	Discrete
<b>S2 CGPA</b>	Discrete
<b>S3 CGPA</b>	Discrete
<b>S4 CGPA</b>	Discrete
<b>S5 CGPA</b>	Discrete
<b>Overall CGPA</b>	Discrete

# Implementation And Testing

During the implementation phase a record of 200 hundred students is collected. The data set is divided into parts first is training and the second is testing. 80% of the data set is used for training and the remaining 20% is used for testing. The training set is used to train the model and the testing set is used to test the accuracy of the model that we have trained. The algorithm that we will be using in this project is Support Vector Machine (SVM) because it has the highest accuracy (81.82%) among other classification techniques.

Support Vector Machine(SVM) is supervised machine learning algorithm that can be used for both regression and classification purpose. The programming skills such as Html and Css will be used to build this project and the tool or IDE that we will be using for writing codes will be PyCharm.

In Order to predict the performance of the student the faculty/teacher has to import following fields listed below :

- 1) Gender
- 2) Name
- 3) Unique id
- 4) Class
- 5) Marks obtained in particular subjects
- 6) Overall percentage

Once these values are Imported than our system will be able to predict the performance of the student.

# CONCLUSION

In conclusion, the Main Objective of this project was to develop a system that predicts the performance of the student. A data mining technique, Support Vector Machine algorithm (SVM) is used in this project to ensure the prediction of the student performance is possible. The main contribution of the student performance analysis system is that it assists the lecturers/faculty to check or monitor the performance of the student. The Project allows lecturers to identify those student who are weak in studies and want extra attention.

# REFERENCES

- [1]. J. Shana, and T. Venkatacalam, “A framework for dynamic Faculty Support System to analyse student course data”, *International Journal of Emerging Technology and Advanced Engineering*, Vol. 2, No. 7, 2012, pp.478-482.
- [2]. Quality Assurance Division (2010). InMinds: Intelligent Mining and Decision Support System. *University of Malaysia Sarawak, UNIMAS* [Online].
- [3]. SPA (2013). What is SPA Standard? *SPA Student Performance Analyser* [Online].
- [4]. V. Kumar, and A. Chadha, “An empirical study of the applications of data mining techniques in higher education”, *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 2, No..3, 2011, pp. 80-84.