**A**

**MINI PROJECT**

**ON**

**STUDENT PERFORMANCE**

**Submitted By**

**Aashish Porwal (160116737084)**

**Kogur Aditya(160116737085)**

**Under the guidance of**

**Mrs Y.Gnyana Deepa**

**Asst. Professor, Dept. of IT, CBIT.**



**DEPARTMENT OF INFORMATION TECHNOLOGY**

**CHAITANYA BHARATHI   INSTITUTE OF TECHNOLOGY(A)**

**(Affiliated to Osmania University; Autonomous under UGC,Accredited by**

**NBA (AICTE) and NAAC-A grade(UGC), ISO 9001:2015 certified Institute)**

**Website:** [**www.cbit.ac.in**](http://www.cbit.ac.in)**, www.cbitworld.com**

**Gandipet,HYDERABAD – 500 075**

**2018-2019**

**CHAITANYA BHARATHI   INSTITUTE OF TECHNOLOGY(A)**

DEPARTMENT OF INFORMATION TECHNOLOGY

**(Affiliated to Osmania University)**

**GANDIPET, HYDERABAD – 500 075**



**CERTIFICATE**

This is to certify that the project work titled “**STUDENT PERFORMANCE**” submitted to **CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY,** in partial fulfilment of the requirements for the award of the completion of 6th semester of B.E. in Information Technology, during the academic year 2018-19, is a record of original work done by

**Aashish Porwal(160116737084)** and **Kogur Aditya(160116737085)** during the period of study in Dept. of IT, CBIT, HYDERABAD, under our supervision and guidance.

**PROJECT GUIDE  HEAD OF DEPARTMENT**

**Ms.Y.Gnyana Deepa        Dr. Suresh Pabboju**

Asst. Professor,       Professor & Head

Dept. of IT, CBIT        Dept. of Information Technology

       CBIT, HYDERABAD

**DECLARATION**

This is to certify that the work reported in the present report titled “**STUDENT PERFORMANCE”** is a record of work done by us in the Department of Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad.

No part of the report is copied from books / journals and wherever the portion is taken, the same has been duly referred. The reported results are based on the project work done entirely by us and not copied from any other source.

Aashish Porwal (160116737084)

            Kogur Aditya (160116737085)

**ACKNOWLEDGEMENT**

It is our privilege to acknowledge with deep sense of gratitude and devotion for keen personal interest and invaluable guidance rendered by our Project Guide   **Ms. Y.Gynana Deepa**, Assistant Professor, Department of Information Technology, Chaitanya Bharathi Institute of Technology.

Our respects and regards to **Dr. Suresh Pabboju**, Professor and Head, Department of Information Technology, Chaitanya Bharathi Institute of Technology, for his invaluable suggestions that helped us in successful completion of the project.

We are grateful to our Principal **Dr.P.Ravinder Reddy,** Chaitanya Bharathi Institute of Technology, for his cooperation and encouragement.

Finally, we also thank all the staff members, faculty of Dept. of IT, CBIT, our friends, and all our family members who with their valuable suggestions and support, directly or indirectly helped us in completing this project work.

Aashish Porwal (160116737084)

            Kogur Aditya (16011673085)

**ABSTRACT**

Student performance is an essential part in higher learning institutions. This is because one of criteria for a high quality institution is based on its excellent record of academic achievemets.

Almost every college has their own management system to manage the student records and keep an eye on their progress but coming to students point of view it is becoming difficult for them to predict their results at the end of the course so our project aims at providing the overall predictable performance at the end of the course based on different social , sports and academic attributes using Neural Network. Or Classification methods.

Therefore, a systematical review on predicting student performance  is proposed to improve students achievements. The main objective of this to provide an overview on the datamining techniques and Machine Learning techniques that have been used to predict students performance.

**CONTENTS**

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Page No. | |
| Abstract | | V | |
| Contents | | Vi | |
| **1.INTRODUCTION** | | 1 | |
| * 1. Introduction And Motivation | | 1 | |
| 1.2 Existing System | | 1 | |
| 1.3 Problem Associated with Existing System  1.4 Objectives  1.5 Proposed System and  advantages | | 2  2  3 | |
| **2.REQUIREMENT ANALYSIS** | | 5 | |
| 2.1 Hardware Requirement | | 5 | |
| 2.1 software Requirement | 5 | |
| **3.DESIGN AND PLANNING** | | 6 | |
| 3.1 Process Model | | 7 | |
| **4. IMPLEMENTATION** | | 8 | |
| **5. SYSTEM TESTING** | | 9 | |
| 5.1 Inputs  5.2 Expected Output and the  Result | | 9  9 | |
| **6.RESULTS** | | 10 | |
| **7.CONCLUSION**  **8.FUTURE ENHANCEMENTS** | | 11  12 | |
| **9. REFERENCES** | | 13 | |
|  | |  | |
|  | |  | |
|  | |  | |
|  | |  | |
|  | |  | |
|  | |  | |
|  | |  | |
|  | |  | |

**1.INTRODUCTION**

* 1. **Introduction and Motivation**

Having spent the past few months studying quite a bit about data mining and statistical inference, I wanted a more serious and challenging task than simply working and re-working the examples that many books and blogs make use of.

My objective was to build a model that would predict whether or not a student would fail the math course that was being tracked. I focused on failure rates as I believed that metric to be more valuable in terms of flagging struggling students who may need more help.

In this work, we will analyze recent real-world data from two Portuguese secondary schools. Two different sources were used: mark reports and questionnaires. Since the former contained scarce information (i.e. only the grades and number of absences were available), it was complemented with the latter, which allowed the collection of several demographic, social and school related attributes (e.g. student’s age, alcohol consumption, mother’s education). The aim is to predict student achievement and if possible to identify the key variables that affect educational success/failure. The two core classes (i.e. Mathematics and Portuguese) will be modeled under three DM goals:

* 1. **Existing System**

System Analysis is a detailed study of the various operations performed by a system and their relationships within and outside of the system. Here the key question is- what all problems exist in the present system? What must be done to solve the problem? Analysis begins when a user or manager begins a study of the program using existing system.

During analysis, data collected on the various files, decision points and transactions handled by the present system. The commonly used tools in the system are Data Flow Diagram, interviews, etc. Training, experience and common sense are required for collection of relevant information needed to develop the system. The success of the system depends largely on how clearly the problem is defined, thoroughly investigated and properly carried out through the choice of solution. A good analysis model should provide not only the mechanisms of problem understanding but also the frame work of the solution. Thus it should be studied thoroughly by collecting data about the system. Then the proposed system should be analyzed thoroughly in accordance with the needs.

System analysis can be categorized into four parts.

System planning and initial investigation

Information Gathering

Applying analysis tools for structured analysis

Feasibility study

Cost/ Benefit analysis.

In the current system we need to keep a number of records related to the student and want to enter the details of the student and the marks manually. In this system only the teacher or the school authority views the mark of the student and they want to enter the details of the student. This is time consuming and has much cost.

.

**1.3 Problem Associated with Existing System**

a) It takes comparatively more time to process.

b) It requires additional devices for operation.

c) It requires external algorithms and memory.

d) For programming it requires development system.

e) It is not user friendly.

f)It has less storage capacity.

**1.4 Objectives**

My objective was to build a model that would predict whether or not a student would fail the math course that was being tracked. I focused on failure rates as I believed that metric to be more valuable in terms of flagging struggling students who may need more help.

To be able to preemptively assess which students may need the most attention is, in my opinion, an important step to personalized education.

**1.5 Proposed System**

In our proposed system we have the provision for adding the details of the students by themselves. So the overhead of the college authorities and the teachers is become less. Another advantage of the system is that it is very easy to edit the details of the student and delete a student when it found unnecessary. The marks of the student are added in the database and so students can also view the marks whenever they want.

Our proposed system has several advantages

* User friendly interface
* Fast access to database
* Less error
* More Storage Capacity
* Search facility
* Look and Feel Environment
* Quick transaction

All the manual difficulties in managing the student details in a college have been rectified by implementing computerization.

**FEASIBILITY ANALYSIS**

Whatever we think need not be feasible .It is wise to think about the feasibility of any problem we undertake. Feasibility is the study of impact, which happens in the organization by the development of a system. The impact can be either positive or negative. When the positives nominate the negatives, then the system is considered feasible. Here the feasibility study can be performed in two ways such as technical feasibility and Economical Feasibility.

**Technical Feasibility:**

We can strongly says that it is technically feasible, since there will not be much difficulty in getting required resources for the development and maintaining the system as well. All the resources needed for the development of the software as well as the maintenance of the same is available in the organization here we are utilizing the resources which are available already.

**Economical Feasibility**

Development of this application is highly economically feasible .The organization needed not spend much money for the development of t he system already available. The only thing is to be done is making an environment for the development with an effective supervision. I f we are doing so, we can attain the maximum usability of the corresponding resources .Even after the development, the organization will not be in condition to invest more in the organization. Therefore, the system is economically feasible.

**ORGANISATION OF REPORT**

The organization of the report is as follows:

**1.** deals with the Introduction of the project and gives the details about the project in an abstract view.

**2.** deals with the information about requirement details for the project are discussed in brief.

**3.** deals with the design and planning for the project.

**4.** deals with the Implementation part which includes the tools and software’s that are used.

**5.** deals with the Testing of the project and screenshots of the project

**6.** explains the results of the project and screenshots of the project.

**7.** explains the conclusion of the project.

**8.** explains the future enhancements of the project.

**9.** deals with the references.

**2.TECHNOLOGIES**

**2.1 Hardware Requirement**

This project is based on both hardware and software. The hardware requirements areas follows :-

**HARDWARE REQUIREMENTS:**

* System : Pentium IV 2.4 GHz.
* Hard Disk : 40 GB.
* Floppy Drive : 1.44 Mb.
* Monitor : 15 VGA Colour.
* Mouse : Logitech.
* Ram : 512 Mb.

**2.2 Software Requirement**

As explained earlier our project requires two part hardware and software. Hardwareparts are explained above and software requires as follows :-

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows XP.
* Coding Language : Python 3.7
* IDE : Eclipse(PyDev)

PyDev is a plugin that enables Eclipse to be used as a Python IDE (supporting also Jython and IronPython). It uses advanced type inference techniques which allow it to provide things such as code completion and code analysis, besides providing a debugger, interactive console, refactoring, tokens browser, django integration, etc.

**3.DESIGN AND PLANNING**

**3.1 Process Model**

Dataset has a large amount of data that has to be arranged in consistent manner. To enhance existing system the proposed model is designed by collecting data and classifying them based on student performance in a particular domain. The performance is classified as

• Poor

• Average

• Excellent

The system framework is as shown below which provides an efficient analysis on student performance by data collection and prediction.

The steps of SVM(Support Vector Machine) process would be followed. The process starts from data collection and data pre-processing followed by classification model construction and ends with model evaluation and interpretations. Steps in SVM is shown in Fig

1. Information gathering :A predefined dataset is selected from the internet constituting of attributs such as name , sex, g1 score , g2 score , schoolsup,sports,etc.
2. Pre-processing 1. Selection of Attributes: There may be many attributes available but all the attribute are not useful for prediction. So only the necessary attributes are taken into consideration to decide the values. 2. Deciding values for Attributes: It is necessary to decide values for attributes to avoid continuous data. This can be done by using discrete data.
3. Model building In this stage, **Confusion Matrix** has been selected as a classifier under cross validation method. For model construction confusion matrix is used based on attributes selected. In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one.
4. **3.2 Data Flow Diagram**

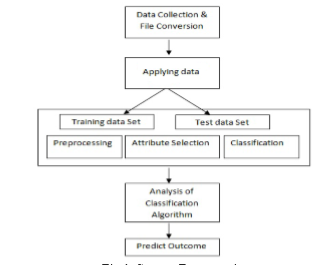
****

Fig 2.1

Fig 2.1 refers to the process in which our algorithm works ,it starts with data collection , followed by training , testing , preprocessing,attribute selection,classification.

**4. IMPLEMENTATION**

1.Load the .CSV files into a dataframe.

2.Preprocessing the data-data cleaning,label encoder, test train and predict

3.Pipeline-to transfer the data

4. Confusion matrix

5.Predict the output

Here are some others factor in classifier output-

•TP Rate: rate of true positives (instances correctly classified as a given class)

• FP Rate: rate of false positives (instances falsely classified as a given class)

• Precision: proportion of instances that are truly of a class divided by the total instances

• Recall: proportion of instances classified as a given class divided by the actual total in that

• class (equivalent to TP rate) F-Measure: A combined measure for precision and recall calculated as 2 \* Precision

• Recall / (Precision + Recall) Sample output to test PDF Combine only P a g e | 11 Prediction of result –First, the file with cases to predict needs to have the same structure that the file used to learn the model.

**5.SYSTEM TESTING**

**5.1 Input**

Values of the attributes in the data set

**5.2 Expected Output**

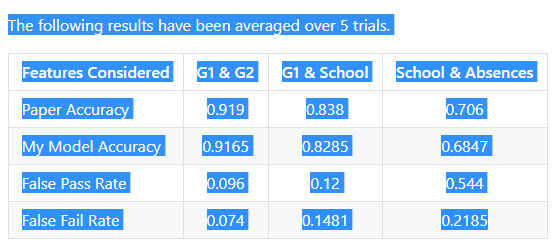
****

Fig 2.2

Fig2.2 represents the output of our experiment , the results are obtauined after 5 trials.

**6.RESULTS**

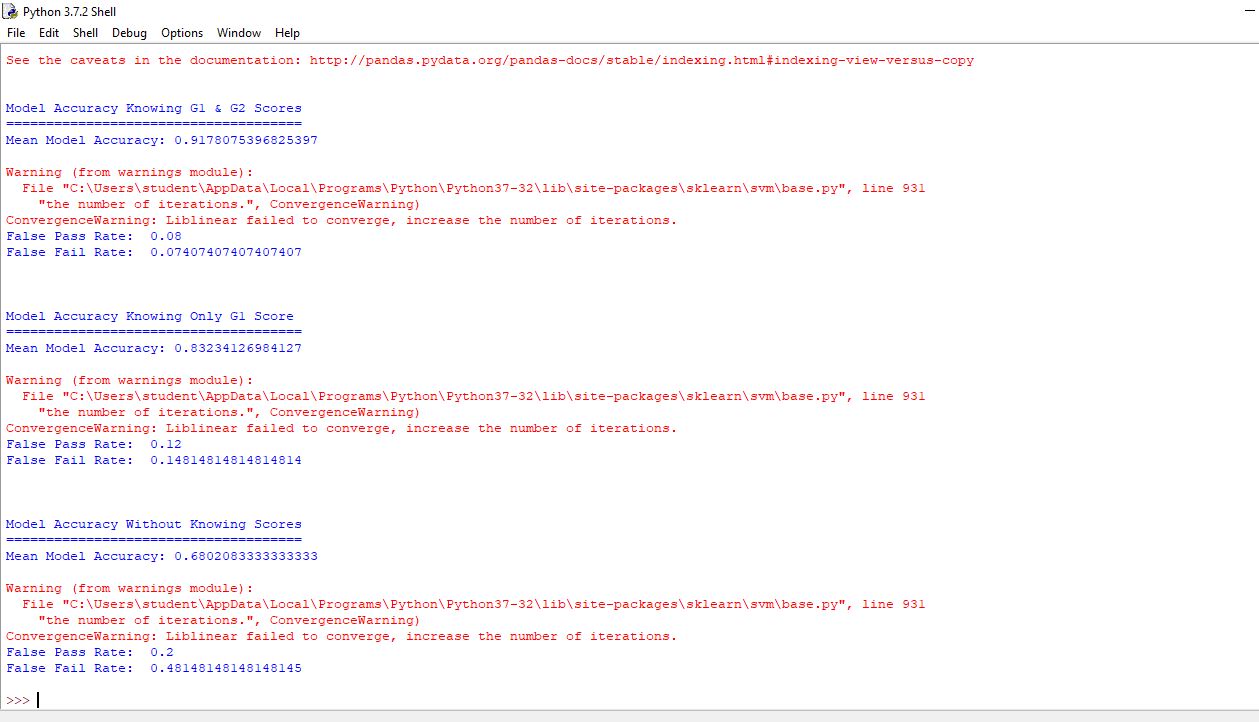
****

Fig 2.3

Fig 2.3 is the screenshot of our working algorithm.

**7.CONCLUSION**

**Conclusion**

Education is a crucial element in our society. Data Mining (DM) techniques, whichallow a high level extraction of knowledge from raw data, offer interesting possibilities for the education domain. In particular, several studies have used DM methods to improve the quality of education and enhance school resource management. In this , we have addressed the prediction of secondary student grades by using past school grades (G1and G2), demographic, social and other school related data. Three different DM goals (i.e. binary/5- level classification and regression) and one DM methods, Support Vector Machines (SVM) were tested. Also, distinct input selections (e.g. with or without past grades) were explored.

The obtained results reveal that it is possible to achieve a high predictive accuracy, provided that the G1 and G2 score are known. Nevertheless, an analysis to knowledge provided by the best predictive models has shown that, in some cases, there are other relevant features, such as: school related (e.g. number of absences, reason to choose school, extra educational school support), demographic (e.g. student’s age, parent’s job and education) and social (e.g. going out with friends, alcohol consumption) variables. Furthermore, we intent to enlarge the experiments to more colleges and college years, in order to enrich the student databases. More research is also needed (e.g. sociological studies) in order to understand why and how some variables (e.g. reason to choose school, parent’s job or alcohol consumption) affect student performance..

**8.FUTURE ENHANCEMENTS**

Educational data mining main focus is to analyse the education system. This paper demonstrates classification method to predict student performance. It also assist in automating the existing manual system by providing the Web Based Information System. It creates connectivity between parents and college. All the stakeholders, faculty and management can get the required information without delay. This whole model can be useful in educational system like MCT Rajiv Gandhi Institute of Technology, Mumbai.

Thus improving their standards and performance.The results of the data mining algorithms for the classification of the students based on the attributes selected reveals that the prediction rates are not uniform. The work can be further extended out by designing the student model analysing records of students extracurricular skills and provide a suggestions on communication and technical skill development by which students can be built in professional aspect of talent.

**9.REFERENCES**

BIBLIOGRAPHY:

[1] Massimo Banzi, 2011, Getting Started with Python, 2nd Edition, UK, Maker Media, Inc.

[2] Parneet Kaura, Manpreet Singhb, Gurpreet Singh Josanc “Classification and Prediction based Data Mining Algorithms to Predict Slow Learners in Education Sector” Science Direct Procedia Computer Science 57 ( 2015 ) 500 – 508 2015 (ICRTC2015).

[3] Baradwaj Brijesh Kumar and Pal Saurabh (2011).Mining Educational Data to Analyze Student Performance. International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6

[4] <http://www3.dsi.uminho.pt/pcortez/student.pdf>

**[5]** <http://www.wikipedia.com>

**[6] http://www.mlrepository.com**