# STUDENT PERFORMANCE ANALYSIS USING MACHINE LEARNING ALGORITHM

# A PROJECT REPORT

***Submitted by***

|  |  |
| --- | --- |
| **M. KAMALESH** | **510619205011** |
| **K.SUDHARSAN** | **510619205040** |
| **R.TAMIL SELVAN** | **510619205044** |

***In partial fulfilment for the award of the degree Of***

**BACHELOR OF TECHNOLOGY IN**

**INFORMATION TECHNOLOGY**

# C.ABDUL HAKEEM COLLEGE OF ENGINEERING AND TECHNOLOGY, MELVISHARAM – 632509

**ANNA UNIVERSITY: CHENNAI - 600025**

# JUNE 2023

**BONAFIDE CERTIFICATE**

Certified that this project report “**STUDENT PERFORMANCE ANALYSIS USING MACHINE LEARNING ALGORITHM**” is the bonafide work of “**M. KAMALESH, K. SUDHARSAN,** and **R. TAMIL SELVAN**” who carried out the project work under my supervision.

# SIGNATURE SIGNATURE

**Dr. S. UMAMAHESWARI, M. Tech., Ph.D., Mr.T.S. KARTHICK, M.E.,**

# PROFESSOR SUPERVISOR

# HEAD OF THE DEPARTMENT ASSISTANT PROFESSOR

|  |  |  |
| --- | --- | --- |
| Department of Information  Technology, | | Department of Information Technology, |
| C. Abdul Hakeem College of Engineering & Technology, | | C. Abdul Hakeem College of Engineering & Technology, |
| Melvisharam, | Melvisharam, | | |
| Vellore-632509. | Vellore-632509. | | |

Submitted for the University Viva-Voce Examination held on .

# INTERNAL EXAMINER EXTERNAL EXAMINER

**ACKNOWLEDGEMENT**

Before we get into thick of things, we would like to add a few genuine words for the people who were the part of this project in numerous ways.

We are very thankful to all people whose content constant support and guidance right from the initial stages has enabled us to complete our project successfully.

We first thank to Almighty to who us given the strength and support during course of this project.

We express our profound gratitude to our honorable chairman **HAJI Dr. S. ZIAUDDIN AHMED sahib, B.A.,** and to our beloved correspondent **HAJI Dr. V.M ABDUL LATHEEF SAHIB, B.E., F.I.E.**

We wish to express our sincere thanks to our respected principal

**Dr. M.SASIKUMAR, Ph.D.,** who has been kind enough to us in all aspects.

We wish to express our heartfelt thanks to our HEAD OF THE DEPARTMENT **Dr. S. UMAMAHESWARI, Ph.D.,** and Internal guide

# Mr.T.S. KARTHIK, M.E.,

And finally, our sincere thanks and gratitude to our parents and friends whose constructive criticism made us to analyse deeper and deeper in our project.

**ABSTRACT**

Behaviour analysis is the science of studying the comportment of a person to establish a specific profile about it. It has been used in psychology and since a few years it has established its presence in Information Technology. With the growth of artificial intelligence, it tends to become the new trend that gives the possibility for applications to be personalized and centered on the user’s needs. Machine Learning is a subcategory of artificial intelligence and has the goal to develop solutions to create automatic methods to make our computers capable of evolving by themselves. The activities and actions of users are analysed and studied to determine rules that can be coordinated to align software applications in parallel with the day to day routine of the person. This thesis is a part of the educational sector in which data mining is used for wide variety of applications to evaluate the performance of the student which focuses on the academic, mental, emotional and social aspects of the students are ignored. The goal of the project is to predict the performance of the student by taking into account various factors like marks, attendance, staff opinion, participation in extracurricular activities, past incidents of ragging and stress. Naïve Bayes and KNN algorithms are used to determine the future performance of the student. A comparative analysis is also performed amongst the two algorithm in terms of time taken by both the algorithms to yield the results.

**சுருக்கம்**

**நடத்தை பகுப்பாய்வு என்பது ஒரு நபரைப் பற்றிய ஒரு குறிப்பிட்ட சுயவிவரத்தை நிறுவுவதற்கு அவரது உள்ளடக்கத்தைப் படிக்கும் அறிவியல் ஆகும். இது உளவியலில் பயன்படுத்தப்பட்டு சில வருடங்களாக தகவல் தொழில்நுட்பத்தில் அதன் இருப்பை நிலைநாட்டியுள்ளது. செயற்கை நுண்ணறிவின் வளர்ச்சியுடன், பயன்பாடுகள் தனிப்பயனாக்கப்பட்ட மற்றும் பயனரின் தேவைகளை மையமாகக் கொண்ட புதிய போக்காக இது மாறுகிறது. இயந்திர கற்றல் என்பது செயற்கை நுண்ணறிவின் ஒரு துணைப்பிரிவாகும், மேலும் நமது கணினிகள் தாமாகவே உருவாகும் திறன் கொண்டதாக மாற்றுவதற்கு தானியங்கி முறைகளை உருவாக்குவதற்கான தீர்வுகளை உருவாக்கும் இலக்கைக் கொண்டுள்ளது. பயனர்களின் செயல்பாடுகள் மற்றும் செயல்கள் பகுப்பாய்வு செய்யப்பட்டு, தினசரி நபரின் அன்றாட வழக்கத்திற்கு இணையாக மென்பொருள் பயன்பாடுகளை ஒருங்கிணைக்கக்கூடிய விதிகளை தீர்மானிக்க ஆய்வு செய்யப்படுகிறது. இந்த ஆய்வறிக்கை கல்வித் துறையின் ஒரு பகுதியாகும், இதில் மாணவர்களின் கல்வி, மன, உணர்ச்சி மற்றும் சமூக அம்சங்களில் கவனம் செலுத்தும் மாணவர்களின் செயல்திறனை மதிப்பிடுவதற்கு பல்வேறு வகையான பயன்பாடுகளுக்கு தரவுச் செயலாக்கம் பயன்படுத்தப்படுகிறது. மதிப்பெண்கள், வருகைப்பதிவு, ஊழியர்களின் கருத்து, சாராத நடவடிக்கைகளில் பங்கேற்பது, கடந்தகால ராகிங் மற்றும் மன அழுத்தம் போன்ற பல்வேறு காரணிகளை கணக்கில் எடுத்துக்கொண்டு மாணவர்களின் செயல்திறனைக் கணிப்பதே திட்டத்தின் குறிக்கோள். நேவ் பேய்ஸ் மற்றும் கேஎன்என் அல்காரிதம்கள் மாணவரின் எதிர்கால செயல்திறனைத் தீர்மானிக்கப் பயன்படுத்தப்படுகின்றன. இரண்டு அல்காரிதங்களுக்கிடையில் ஒரு ஒப்பீட்டு பகுப்பாய்வு இரண்டு வழிமுறைகளும் முடிவுகளை வழங்குவதற்கு.**

## TABLE OF CONTENTS

|  |  |  |  |
| --- | --- | --- | --- |
| CHAPTER NO. |  | TITLE | PAGE NO. |
|  | ABSTRACT | | iii |
|  | ACKNOWLEDGEMENT | | iv |
|  | LIST OF FIGURES | | vii |
|  | LIST OF ABBREVIATIONS | | viii |
| 1. | INTRODUCTION | | 1 |
|  | 1.1 | GENERAL | 1 |
|  | 1.2 | OBJECTIVE | 1 |
|  | 1.3 | EXISTING SYSTEM | 2 |
|  | 1.4 | DISADVANTAGES OF THE EXISTING SYSTEM | 2 |
|  | 1.5 | PROPOSED SYSTEM | 2 |
|  | 1.6 | ADVANTAGES OF PROPOSED SYSTEM | 2 |
| 2 | LITERATURE REVIEW | | 3 |
|  | 2.1 | USE OF DISTRIBUTED DATA MINING TO PREDICT THE PERFORMANCE OF STUDENTS | 3 |
|  | 2.2 | ACADEMIC PERFORMANCE PREDICTION OF  STUDENTS | 3 |
|  | 2.3 | PREDICTING STUDENTS PERFORMANCE USING DATA MINING | 4 |
|  | 2.4 | USE OF DATA MINING TO UPGRADE INDIAN EDUCATION SYSTEM | 4 |
|  | 2.5 | TECHNIQUES AND FACTORS USED IN EDUCATIONAL DATA MINING TO PREDICT STUDENT AMELIORATION | 4 |
| 3 | PROPOSED METHODOLOGY | | 6 |
|  | 3.1 | MODULES | 6 |
|  | 3.2 | SYSTEM WORKFLOW | 7 |
|  | 3.3 | ALGORITHM | 7 |
| 4 | REQUIREMENT ANALYSIS | | 9 |
|  | 4.1 | HARDWARE REQUIREMENTS | 9 |
|  | 4.2 | SOFTWARE REQUIREMENTS | 9 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | 4.3 | FUNCTIONAL REQUIREMENTS | 9 |
|  | 4.4 | NON-FUNCTIONAL REQUIREMENTS | 10 |
| 5 | SYSTEM DESIGN | | 12 |
|  | 5.1 | INPUT DESIGN | 12 |
|  | 5.2 | OUTPUT DESIGN | 12 |
| 6 | SOFTWARE SPECIFICATION | | 14 |
|  | 6.1 | INTRODUCTION | 14 |
|  | 6.2 | APPLICATION OF JAVA | 14 |
|  | 6.3 | FEATURES OF JAVA | 14 |
|  | 6.4 | COLLECTION FRAMEWORK | 16 |
|  | 6.5 | MYSQL | 17 |
| 7 | TEST CASES AND SNAPSHOTS | | 19 |
| 8 | SOFTWARE TESTING | | 24 |
|  | 8.1 | GENERAL | 24 |
|  | 8.2 | TYPES OF TESTING | 25 |
| 9 | CONCLUSION | | 27 |
| 10 | FUTURE ENHANCEMENT | | 28 |
| 11 | REFERENCES | | 29 |
|  | APPENDIX | | 30 |
|  | PAPER PUBLICATION STATUS | | 53 |
|  | PLAGIARISM REPORT | | 54 |

**LIST OF FIGURES**

System Workflow 7

* 1. K Means 8
  2. Compilation of Java 15
  3. Collection Framework 17
  4. Sample Test Cases 19
  5. User Interface 20
  6. User Input-1 20
  7. Output-1… 21
  8. Algorithm Accuracy-1 21
  9. Dataset… 22
  10. User Input-2. 22
  11. Output-2… 23
  12. Algorithm Accuracy-2. 23

## LIST OF ABBREVIATIONS

**KNN**

K-Nearest Neighbor

**JVM**

Java Virtual Machine

**MySQL** My Sequel

**GPU** Graphics Processing Unit

**LAMP** Linux, Apache, MySQL and PHP

## CHAPTER 1 INTRODUCTION

### GENERAL

Many improvements have been made in education using data mining techniques. Data mining is a technique of extracting hidden patterns from huge databases. These concepts and methods of data mining can be applied in various domains like marketing, real estate, customer relationship management, engineering, web mining, etc.

Educational data mining is an emerging sector which is concerned with creating and developing improved methods for extracting knowledge from the data obtained from the educational sector. The data which basically includes the personal information or academic performance of the student can be obtained from the past experiences or from the operational data which resides in the databases of educational institutes. Furthermore, the data can also be obtained from e-learning database systems which have a huge amount of data and information used by most institutes.

Many techniques have been used for proper implementation of data mining concepts such as Naïve Bayes and KNN. Use of these techniques provides different kinds of knowledge which can be discovered using classification and clustering. By using these we can extract knowledge that describes student’s performance in examination and all their detailed information. From the large amount of data obtained, the data is firstly sorted, cluster analysis is to classify the raw data. Clustering is a bunch of physical or abstract objects which are divided and regrouped together on the basis of similarity.

### OBJECTIVE

The main aim of this project is to improvise the student performance in studies based on academic as well as non- academic factors. Education is an essential factor which leads to the development of a country. Developing new methods to obtain useful information form educational database and find a trend amongst students to provide them suitable support a

guidance can improve the educational scenario of the country. The analysis and summarization of data ranges across various dimensions. This motivated us to work on student dataset analysis. The data collection, categorization and classification is thereafter performed manually.

### EXISTING SYSTEM

As of now, existing system takes only academic performance i.e. the result available from the past examinations to predict the future performance of a student.

### DISADVANTAGES OF THE EXISTING SYSTEM

The major disadvantage of the existing system is the ignorance of various non-academic factors which could adversely affect the student’s performance. We propose a system which would take into consideration all the factors which could hamper the performance of the student in the future.

### PROPOSED SYSTEM

The main aim is to develop a trust model using data mining techniques, which mines required information, so that the present education system can adopt this tool. The proposed system uses educational data mining techniques to evaluate performance of a student and identify undesirable behavior. In educational sector, data mining is used for wide variety of applications such as performance of the students like mark, attendance, staff opinion, extracurricular activities, ragging and stress. The data mining techniques used for identifying the performance of the student using K-means and KNN algorithms.

### ADVANTAGES OF PROPOSED SYSTEM

Educational database contain the useful information for evaluating the students. The data mining techniques are helpful in classifying educational database and evaluating the performance of a student.

## CHAPTER 2 LITERATURE REVIEW

### USE OF DISTRIBUTED DATA MINING TO PREDICT THE PERFORMANCE OF STUDENTS

The performance of students in higher education in India is a turning point in the academics for all students for their brightest career. As the importance of education is being realized, the amount of data stored in the educational database is growing exponentially. This educational data is distributed across various nodes in a distributed database. This database contains information which if used effectively can help to improve a student’s performance. The major techniques which are used in data mining across various fields are namely classification and prediction. The basic purpose of classification is to predict the performance of a student in a distributed environment. Data mining is implemented to analyze the available data and extract useful information which can support decision making. The results of models which are present at local level make it difficult to extract knowledge which can be used at global level. Therefore, to support decision making at this area, it is important to generalize the information contained in those models, specific classifier method can be used to generalize these rules for global model.

### ACADEMIC PERFORMANCE PREDICTION OF STUDENTS

The field of engineering observes a relatively high attrition rate worldwide. According to statistics, 35% of the first-year students in various engineering programs do not make it to the second year. Many of the students either drop out or fail in their second or third year of engineering raising the percentage even higher. The major purpose of this investigation is to identify the factors that serve as good indicators to determine whether a student will drop out or fail the program. In order to display early warning , principal component analysis is used on the first-year engineering student academic records. These performance predictors, if identified, can then be used effectively to formulate corrective action plans to improve the attrition rate.

### PREDICTING STUDENTS PERFORMANCE USING DATA MINING

This work proposes a novel approach - personalized forecasting - to consider the successive impact in foreseeing understudy execution (PSP). Rather than utilizing every single verifiable datum as different strategies in PSP, the proposed techniques just utilize the data of the individual understudies for gauging his/her very own execution. Additionally, these techniques likewise encode the "understudy impact" (for example how great/shrewd an understudy is, in playing out the assignments) and "errand impact" (for example how troublesome/simple the errand is) into the models. Trial results demonstrate that the proposed techniques perform pleasantly and a lot quicker than the other best in class strategies in PSP.

### USE OF DATA MINING TO UPGRADE INDIAN EDUCATION SYSTEM

Education is the backbone of all developing countries. Improvisation of the education system leads to the upgradation of the world ranking of the country. One of the major challenges faced by the education system facing is to predict the behavior of students from large database. This paper focus on upgrading Indian education system by using one of the techniques in data mining namely clustering. Cluster analysis distributes the available data into some meaningful groups. The performances of the students are then classified into different patterns as normal, average and below average. In this paper we tend to analyze student’s data on various factors through newly proposed UCAM (Unique clustering with Affinity Measures) clustering algorithm.

### TECHNIQUES AND FACTORS USED IN EDUCATIONAL DATA MINING TO PREDICT STUDENT AMERLIORATION

Educational Data Mining (EDM) is an interdisciplinary ingenuous research area that improves the methods to investigate data arising in educational fields. The methodology used by EDM is to examine academic data in order to study educational questions. As a result, it provides intrinsic knowledge of teaching and learning process for effective education planning. This paper conducts a comprehensive study on the ongoing and significant studies put through in this field to date. The examination centers around strategies for investigating instructive information to create models for improving scholarly

exhibitions and improving institutional sufficiency. The results of these studies give insight into techniques for ameliorating instructive procedures, presaging student performance, compare the precision of data mining algorithms, and demonstrate the development of open source implements.

## CHAPTER 3 PROPOSED METHODOLOGY

### MODULES

**DATA COLLECTION**

In this module, student data’s will be collected from the college. Student’s data will include the following components like mark, attendance, staff opinion, extracurricular activities, ragging and stress. This can also be extracted by circulating a form amongst the students and gaining their feedbacks.

### PRE PROCESSING

Information pre-handling is done to evacuate the fragmented uproarious and conflicting information. Information must be pre-prepared before utilizing in highlight choice assignment.

### CLASSIFICATION MODULE

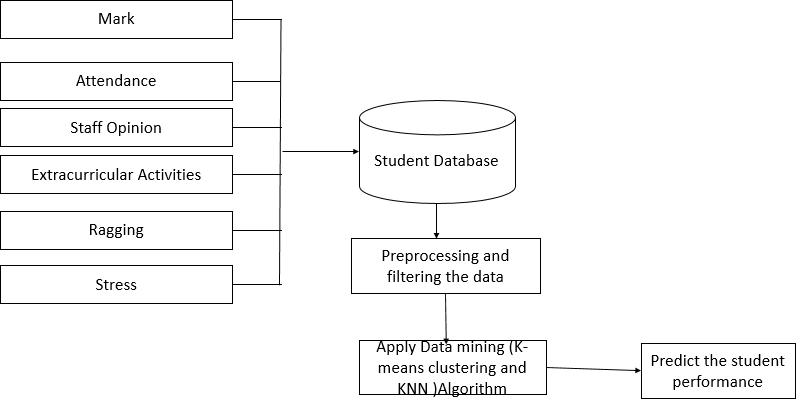
The data mining techniques used for identifying the performance of the student using K- means and KNN algorithms. These two algorithms identify and analyses the performance of the student.

### PREDICTION

In this module, the student’s performance is predicted based upon student mark,

attendance, staff opinion, extracurricular activities, ragging and Stress.

### SYSTEM WORKFLOW



**Figure 3.1:** System Workflow

### ALGORITHM

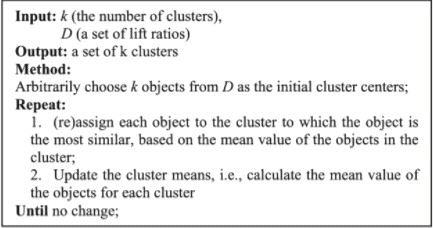
**EANS CLUSTERING ALGORITHM:**

1. Partitional clustering approach
2. Each cluster is associated with a centroid (center point)
3. Each point is assigned to the cluster with the closest centroid 4 Number of clusters K must be specified

### NAIVE BAYES ALGORITHM:

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable.

An advantage of naive Bayes is that it only requires a small number of training data to estimate the parameters necessary for classification.



**Figure 3.2:** K-Means

### EAREST NEIGHBOR (KNN) CLASSIFICATION METHOD:

K-NN is a sort of event - based learning, or languid realizing, where the capacity is just approximated locally and all calculation is conceded until characterization. The K-NN calculation is one of the most straightforward of all AI calculations. The neighbors are gotten from a lot of articles for which the class (for k-NN order) or the item property estimation (for k-NN relapse) is known.

STEP 1: BEGIN

STEP 2: Input: D = {(x1, c1), . . . , (xN , cN )}

STEP 3: x = (x1. . . xn) new instance to be classified

STEP 4: FOR each labelled instance (xi, ci) calculate d (xi, x) STEP 5: Order d (xi, x) from lowest to highest, (i = 1. . . N) STEP 6: Select the K nearest instances to x: Dkx

STEP 7: Assign to x the most frequent class in Dkx STEP 8: END

## CHAPTER 4 REQUIREMENT ANALYSIS

Requirement analysis is the process of determining the requirements of a user for a new modified product. It encompasses the tasks that determine the need for analyzing, documenting, approving and overseeing programming or framework prerequisites. The requirements should be documentable, actionable, measurable, testable and traceable related to identified business needs or opportunities also, characterize to a dimension of detail, adequate for framework plan.

### HARDWARE REQUIREMENTS

System – Intel i5

Speed – 1.8GHZ

Hard disk - 120GB

Monitor – 13.3 inch RAM – 8GB

### SOFTWARE REQUIREMENTS

Operating System – macOS

Mojave Coding language –Python

IDE – PyCharm ,Jupiter notebook Database – MYSQL

### FUNCTIONAL REQUIREMENTS

It is a technical specification requirement for the software products. It is the first step in the requirement analysis process which lists the requirements of particular software systems including functional, performance and security requirements. The function of the system depends mainly on the quality hardware used to run the software with given functionality.

### Usability

It specifies how easy the system must be use. It is easy to ask queries in any format which is short or long, porter stemming algorithm stimulates the desired response for user.

### Robustness

It refers to a program that performs well not only under ordinary conditions but also under unusual conditions. It is the ability of the user to cope with errors for irrelevant queries during execution.

### Security

The state of providing protected access to resource is security. The system provides good security and unauthorized users cannot access the system there by providing high security. **Reliability**

It is the probability of how often the software fails. The measurement is expressed in MTBF (Mean Time between Failures). The requirement is needed in order to ensure that the processes work correctly and completely without being aborted. It can handle any load and survive and survive and even capable of working around any failure.

### Compatibility

It is supported by version above all web browsers. Using any web servers like localhost makes the system real-time experience.

### Flexibility

The flexibility of the project is provided in such a way that is has the ability to run on different environments being executed by different users.

### Safety

Safety is a measure taken to prevent trouble. Every query is processed in a secured manner without letting others to know one’s personal information.

### NON- FUNCTIONAL REQUIREMENTS

**Portability**

It is the usability of the same software in different environments. The project can be run in any operating system.

### Performance

These requirements determine the resources required, time interval, throughput and everything that deals with the performance of the system.

### Accuracy

The result of the requesting query is very accurate and high speed of retrieving information. The degree of security provided by the system is high and effective.

### Maintainability

Project is simple as further updates can be easily done without affecting its stability. Maintainability basically defines that how easy it is to maintain the system. It means that how easy it is to maintain the system, analyse, change and test the application. Maintainability of this project is simple as further updates can be easily done without affecting its stability.

## CHAPTER 5 SYSTEM DESIGN

### INPUT DESIGN

The input design is the connection between the information system and the client. It involves the developing specifications and strategies for data preparation and those steps are important to put transaction data in to a usable form for processing can be achieved by investigating the computer to read data from a written or printed document or on the other hand it can happen by having individuals entering the information straightforwardly into the framework. The main purpose of design is to focus on controlling the amount of required input, controlling the errors, avoiding delay, avoiding unnecessary steps and keeping the process straightforward. The input is designed in such a way to provide security and ease of use while retaining the privacy. Input design considered the following things:

* + - The data to be given as input
    - The arrangement of data
    - The method to guide the operating personnel
    - Way to provide input validation
    - Steps to be followed in case of error.

### OUTPUT DESIGN

A quality output is the one that meets the necessities of the user and presents the data in a user friendly format. In any system results of the processing unit are communicated to the user and to different system through outputs. The output style helps to determine how the data would be displayed to the immediate user. Economical and intelligent output style improves the relationship of the system with the user and support the decision making of the user.

The output form of an information system should focus on accomplishing one or more of the following objectives.

* + - Convey detailed information about past activities, current status or projections of the future.
    - Notify important events, opportunities, problems, or warnings.
    - Trigger an action.
    - Confirm an action.

## CHAPTER 6

**SOFTWARE SPECIFICATION**

**Python**

**Python** is a very popular general-purpose interpreted, interactive, object-oriented, and high-level programming language. Python is dynamically-typed and garbage-collected programming language. It was created by Guido van Rossum during 1985- 1990. Like Perl, Python source code is also available under the GNU General Public License

**Python** is a MUST for students and working professionals to become a great Software Engineer specially when they are working in Web Development Domain. I will list down some of the key advantages of learning Python:

* **Python is Interpreted** − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
* **Python is Interactive** − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
* **Python is Object-Oriented** − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
* **Python is a Beginner's Language** − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

## 4.1 Characteristics of Python

Following are important characteristics of **Python Programming** −

* It supports functional and structured programming methods as well as OOP.
* It can be used as a scripting language or can be compiled to byte-code for building large applications.
* It provides very high-level dynamic data types and supports dynamic type checking
* It supports automatic garbage collection.
* It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

## Applications of Python

The latest release of Python is 3.x. As mentioned before, Python is one of the most widely used language over the web. I'm going to list few of them here:

* **Easy-to-learn** − Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
* **Easy-to-read** − Python code is more clearly defined and visible to the eyes.
* **Easy-to-maintain** − Python's source code is fairly easy-to-maintain.
* **A broad standard library** − Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
* **Interactive Mode** − Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
* **Portable** − Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
* **Extendable** − You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
* **Databases** − Python provides interfaces to all major commercial databases.
* **GUI Programming** − Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
* **Scalable** − Python provides a better structure and support for large programs than shell scripting.

# 4.2 Python – Functions

A function is a block of organized, reusable code that is used to perform a single, related action. Functions provide better modularity for your application and a high degree of code reusing.

As you already know, Python gives you many built-in functions like print(), etc. but you can also create your own functions. These functions are called *user-defined functions.*

## Defining a Function

You can define functions to provide the required functionality. Here are simple rules to define a function in Python.

* Function blocks begin with the keyword **def** followed by the function name and parentheses ( ( ) ).
* Any input parameters or arguments should be placed within these parentheses. You can also define parameters inside these parentheses.
* The first statement of a function can be an optional statement - the documentation string of the function or *docstring*.
* The code block within every function starts with a colon (:) and is indented.
* The statement return [expression] exits a function, optionally passing back an expression to the caller. A return statement with no arguments is the same as return None.

## Syntax

def functionname( parameters ):

"function\_docstring"

function\_suite

return [expression]

By default, parameters have a positional behavior and you need to inform them in the same order that they were defined.

## Example

The following function takes a string as input parameter and prints it on standard screen.

def printme( str ):

"This prints a passed string into this function"

print str

return

## Calling a Function

Defining a function only gives it a name, specifies the parameters that are to be included in the function and structures the blocks of code.

Once the basic structure of a function is finalized, you can execute it by calling it from another function or directly from the Python prompt. Following is the example to call printme() function −

#!/usr/bin/python

# Function definition is here

def printme( str ):

"This prints a passed string into this function"

print str

return;

# Now you can call printme function

printme("I'm first call to user defined function!")

printme("Again second call to the same function")

When the above code is executed, it produces the following result −

I'm first call to user defined function!

Again second call to the same function

## Pass by reference vs value

All parameters (arguments) in the Python language are passed by reference. It means if you change what a parameter refers to within a function, the change also reflects back in the calling function.

**4.3 NUMPY LIBRARY**

NumPy is a Python library used for working with arrays.

It also has functions for working in domain of linear algebra, fourier transform, and matrices.

NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.

NumPy stands for Numerical Python.

## Why Use NumPy?

In Python we have lists that serve the purpose of arrays, but they are slow to process.

NumPy aims to provide an array object that is up to 50x faster than traditional Python lists.

The array object in NumPy is called ndarray, it provides a lot of supporting functions that make working with ndarray very easy.

Arrays are very frequently used in data science, where speed and resources are very important.

**Data Science:** is a branch of computer science where we study how to store, use and analyze data for deriving information from it.

## Why is NumPy Faster Than Lists?

NumPy arrays are stored at one continuous place in memory unlike lists, so processes can access and manipulate them very efficiently.

This behavior is called locality of reference in computer science.

This is the main reason why NumPy is faster than lists. Also it is optimized to work with latest CPU architectures.

## Which Language is NumPy written in?

NumPy is a Python library and is written partially in Python, but most of the parts that require fast computation are written in C or C++.

# 4.4 Machine Learning with Python

Machine Learning (ML) is basically that field of computer science with the help of which computer systems can provide sense to data in much the same way as human beings do. In simple words, ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method. The key focus of ML is to allow computer systems to learn from experience without being explicitly programmed or human intervention.

## Need for Machine Learning

Human beings, at this moment, are the most intelligent and advanced species on earth because they can think, evaluate and solve complex problems. On the other side, AI is still in its initial stage and haven’t surpassed human intelligence in many aspects. Then the question is that what is the need to make machine learn? The most suitable reason for doing this is, “to make decisions, based on data, with efficiency and scale”.

Lately, organizations are investing heavily in newer technologies like Artificial Intelligence, Machine Learning and Deep Learning to get the key information from data to perform several real-world tasks and solve problems. We can call it data-driven decisions taken by machines, particularly to automate the process. These data-driven decisions can be used, instead of using programing logic, in the problems that cannot be programmed inherently. The fact is that we can’t do without human intelligence, but other aspect is that we all need to solve real-world problems with efficiency at a huge scale. That is why the need for machine learning arises.

## Why & When to Make Machines Learn?

We have already discussed the need for machine learning, but another question arises that in what scenarios we must make the machine learn? There can be several circumstances where we need machines to take data-driven decisions with efficiency and at a huge scale. The followings are some of such circumstances where making machines learn would be more effective −

### Lack of human expertise

The very first scenario in which we want a machine to learn and take data-driven decisions, can be the domain where there is a lack of human expertise. The examples can be navigations in unknown territories or spatial planets.

### Dynamic scenarios

There are some scenarios which are dynamic in nature i.e. they keep changing over time. In case of these scenarios and behaviors, we want a machine to learn and take data-driven decisions. Some of the examples can be network connectivity and availability of infrastructure in an organization.

**4.5 PYTHON MYSQL**

Python can be used in database applications.

One of the most popular databases is MySQL.

## MySQL Database

To be able to experiment with the code examples in this tutorial, you should have MySQL installed on your computer.

You can download a MySQL database at <https://www.mysql.com/downloads/>.

## Install MySQL Driver

Python needs a MySQL driver to access the MySQL database.

In this tutorial we will use the driver "MySQL Connector".

We recommend that you use PIP to install "MySQL Connector".

PIP is most likely already installed in your Python environment.

Navigate your command line to the location of PIP, and type the following:

Download and install "MySQL Connector":

C:\Users\Your Name\AppData\Local\Programs\Python\Python36-32\Scripts>python -m pip install mysql-connector-python

**4.6 Wamp Server**

**Windows, Apache, MySQL and PHP**is commonly abbreviated as **WAMP.** Some people may confuse with **LAMP**but the only difference between the two is their operating systems. In case of LAMP, L stands for Linux. Setting up a server included the installation of all the software listed in the abbreviation. Another version is **MAMP,** which is for Mac.

**Why should we set up a WAMP server?**

Everything listed in the abbreviation can be downloaded separately but it takes time to configure each of them. In case of WAMP, the time taken to this is much less than it comes as a whole package. It is used in web development to have a safe experience in testing features.

**Steps to set up a server:**

1. **Download:**Go to https://www.wampserver.com/en/ and install the version that is appropriate for your system. After installation, run the installer. During this part, you can change your default browser if you want. You can go ahead with the default options or add more as per your liking. If you want to use it on multiple machines, installing it on a flash drive would do the job.

**2. Configuration:**Now, we have to configure the WAMP contents we installed. Once installed, you will mostly get a notification from the firewall asking whether the newly installed software should get the permission to use your network. Give it the permission and next find the option in your hidden taskbar icons or in the windows start menu. The color of the symbol corresponds to the status the server is in:

**Red-**Can mean the WAMP server is temporarily deactivated or there is some sort of hindrance that is not allowing it to work

**Orange-**Can mean it is idle or, like red, there is something that didn’t get installed properly.

**Green-**The server is active and ready to use

### MYSQL

MySQL, formally, yet additionally called "My Sequel" is the world's most generally utilized open-source social database the board framework (RDBMS) that keeps running as a server giving multi-client access to various databases, however SQLite likely has progressively complete inserted arrangements. The SQL expression represents Structured Query Language.

The MySQL improvement venture has made its source code accessible under the terms of the GNU General Public License, just as under an assortment of exclusive understandings. MySQL was possessed and supported by a solitary revenue driven firm, the Swedish organization MySQL AB, presently claimed by Oracle Corporation. MySQL is a prominent decision of database for use in web applications, and is a focal segment of the generally utilized LAMP open source web application programming stack (and other 'AMP' stacks). Light is an abbreviation for "Linux, Apache, MySQL, Perl/PHP/Python."

Free software open source extends that require a full-included database the board framework frequently use MySQL.

For business use, a few paid releases are accessible, and offer extra usefulness. Applications which use MySQL databases include: TYPO3, MODx, Joomla, WordPress, phpBB, MyBB, Drupal and other programming. MySQL is likewise utilized in some prominent, extensive scale sites, including Wikipedia, Google (however not for quests), Facebook, Twitter, Flicker and YouTube.

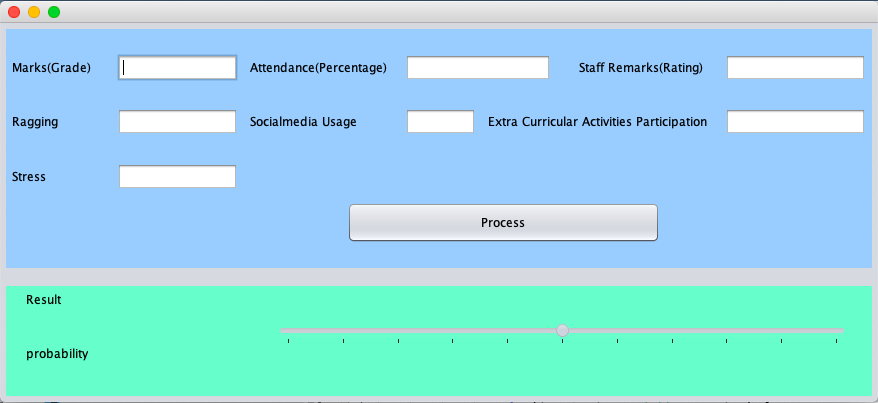
## CHAPTER 7

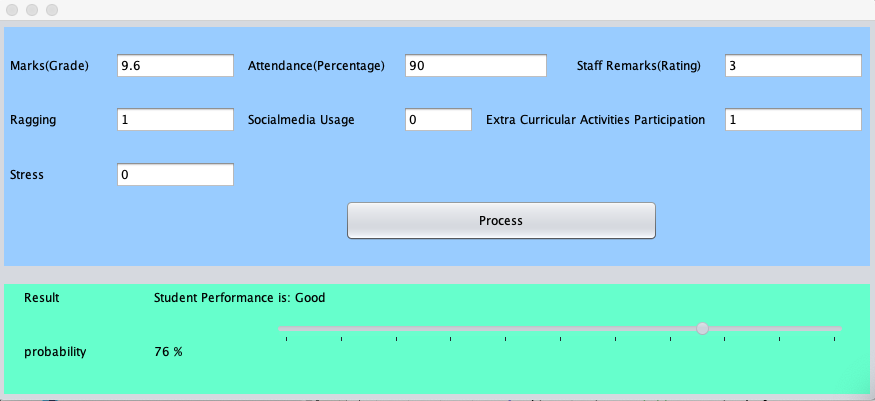
**TEST CASES AND SNAPSHOTS**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Marks (Grade) | Attendance (Percentage) | Staff remarks (Ratings) | Ragg ing | Social- media  Usage | Extra Curricular Activities participation | Str ess | Res ult |
| 9.6 | 90 | 3 | 1 | 0 | 1 | 1 | Yes |
| 6.5 | 78 | 2 | 0 | 1 | 1 | 0 | Yes |
| 5.3 | 60 | 1 | 1 | 1 | 0 | 1 | No |
| 7 | 75 | 2 | 1 | 0 | 0 | 0 | Yes |
| 8 | 81 | 3 | 0 | 1 | 1 | 0 | Yes |

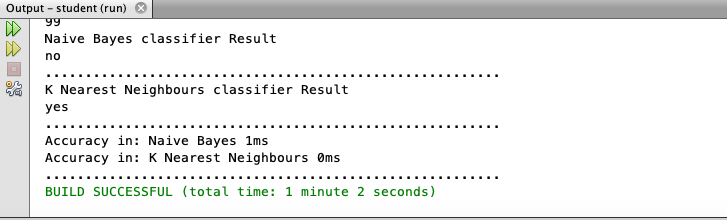
**Figure 7.1**: Sample Test Cases

* Marks (Grades) out of 10
* Attendance percentage out of 100
* Staff Remarks out of 5
* Extra-curricular activities participation [ Yes (1) or No (0) ]
* Stress [ Yes (1) or No (0) ]
* Ragging (Emotional Factors) [ Yes (1) or No (0) ]

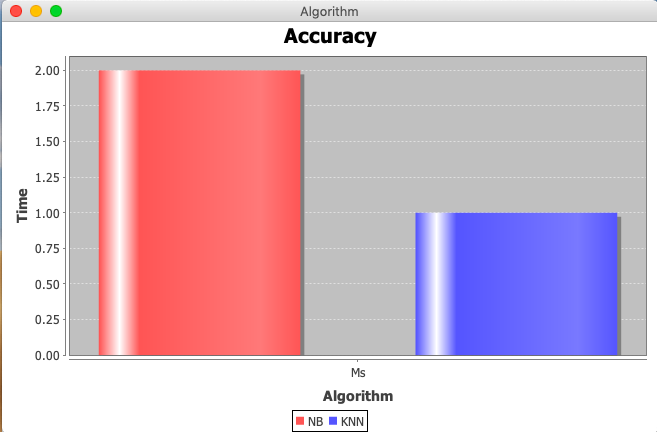
**Figure 7.2**: User Interface



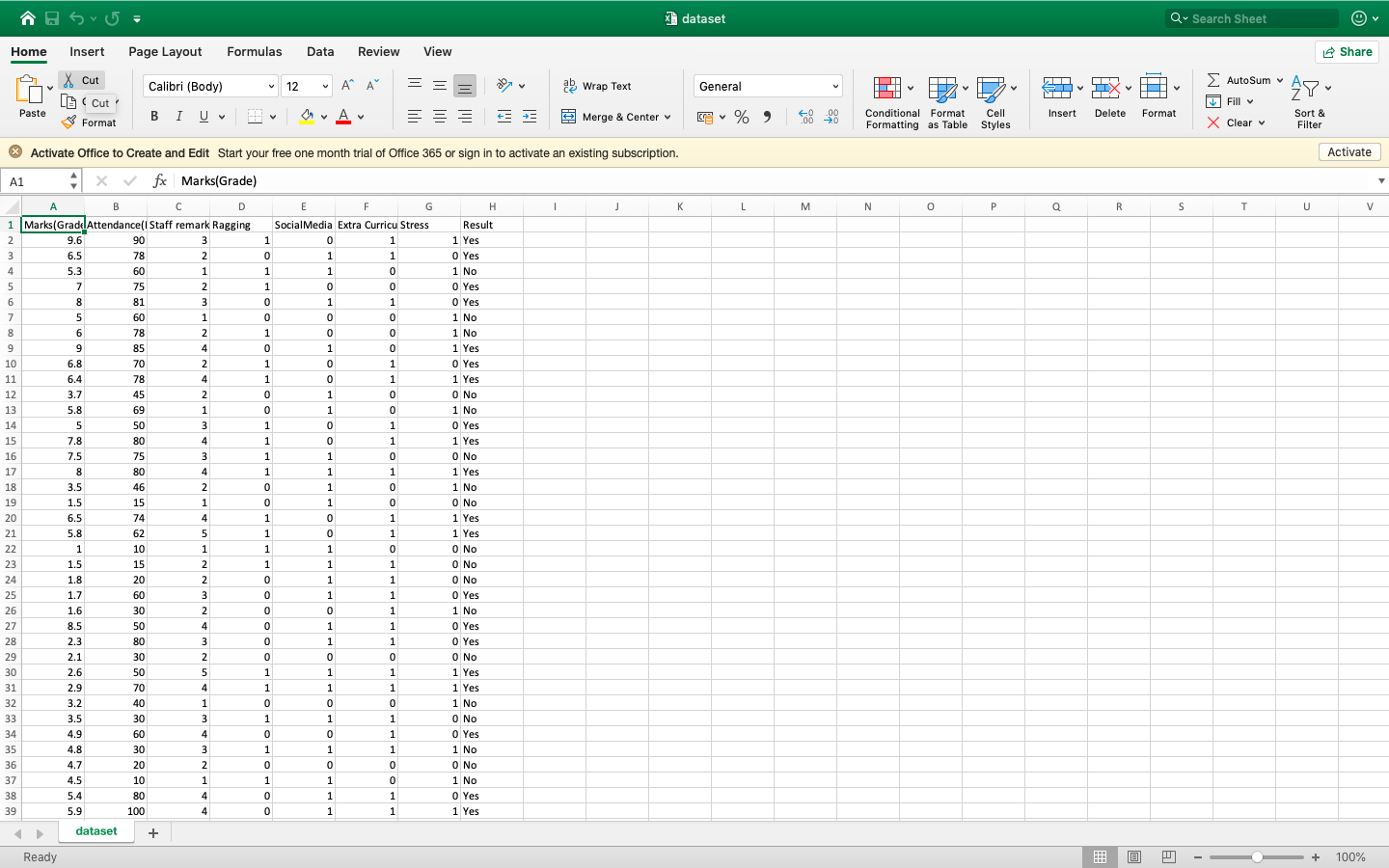
**Figure 7.3:** User input-1



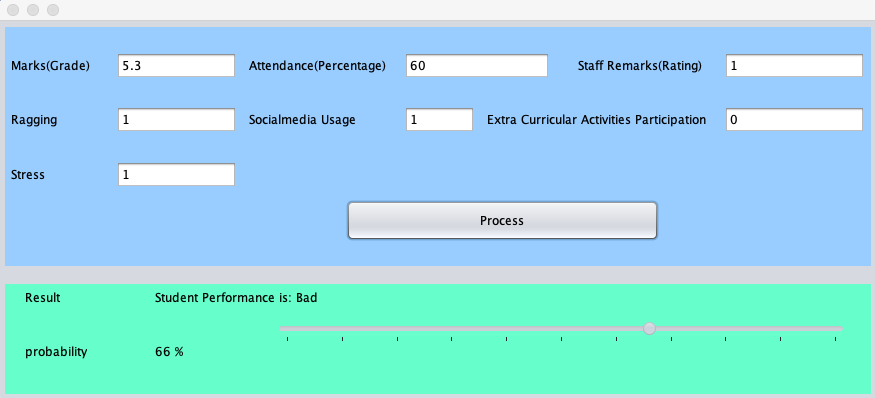
**Figure 7.4:** Output-1



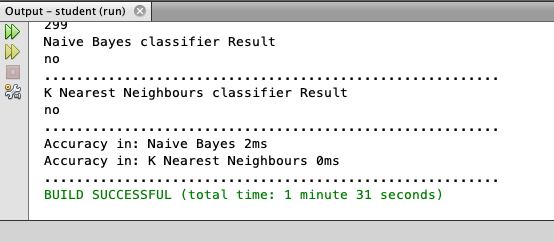
**Figure 7.5:** Algorithm Accuracy-1



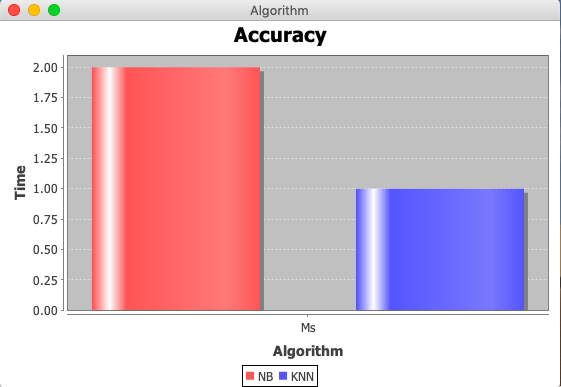
**Figure 7.6:** Dataset



**Figure 7.7**: User Input-2



**Figure 7.8**: output-2



**Figure 7.9:** Algorithm Accuracy-2

## CHAPTER 8

**SOFTWARE TESTING**

### GENERAL

**Test Plan**

Software testing is the method of analysis of a software package item to notice variations between given input and expected output. It is also done to assess the feature of a software package item. Testing assesses the standard of the merchandise. Software package testing may be a method that ought to be done throughout the event method. Alternatively, software package testing may be a verification and validation method.

### Verification

Verification is the method to form certain that the merchandise satisfies the conditions obligatory at the beginning of the event section. In other words, it is a method to form certain that the merchandise behaves as per the approach we wish it to.

### Validation

Validation is the process of ensuring that product development has been done in alignment with the requirements as specified by the customer.

### Basics of Software Testing

Software testing basics can be classified as:

* Black Box Testing
* White Box Testing

### Black Box Testing

Black Box Testing technique ignores the inner mechanism of the system and focuses on the output generated against any input and execution of the system. It is also known by the term Functional Testing.

### White Box Testing

White Box Testing technique takes into consideration the interior mechanism of a system. It is additionally known as Structural Testing and Glass Box Testing.

Black Box testing is usually used for validation and White Box Testing is usually used for verification.

### TYPES OF TESTING

The different types of testing are:

* + - Unit Testing
    - Integration Testing
    - Functional testing
    - System Testing
    - Stress Testing
    - Performance Testing
    - Usability Testing
    - Acceptance Testing
    - Regression Testing
    - Beta Testing

### Unit Testing

Unit testing is performed on an individual unit or a cluster of connected units. It belongs to the category of white box testing. It is usually done to check whether the unit implemented by the programmer generates expected output against the given input.

### Integration Testing

Integration testing is performed on a combined group of components to test the produced output. It further checks the interaction between the software and hardware components if they are related to each other. It could be categorized under both white box testing and black box testing.

### Functional Testing

It is the testing to ensure that the functionalities of the system work as per those specifies in the requirements. It falls in the category of Black box testing.

### System Testing

System Testing technique is done to ensure that the system works in different environments, for eg. Operating systems. It involves the full implementation of the system and environment. It belongs to the category of Black Box testing.

### Stress Testing

Stress testing is done to check the response of the system under unfavorable conditions. It involves extending limits beyond specified in the testing process. It belongs under Black Box testing.

### Performance Testing

Performance testing technique tests the speed and effectiveness of the system and is done to ensure that the system responds within the time specified in the requirements. It is also classified under Black Box Testing.

### Usability Testing

Usability testing is performed to test the system from the user’s perspective. This involves the testing of whether the GUI is user-friendly, the ease of understanding of the user, and checking the proficiency of performance by the user, and whether the design is appealing to the users or not. It also falls under the class of Black box testing.

### Acceptance Testing

Acceptance testing technique is performed by the customers to ensure the final product meets the user’s expectations and requirements. It belongs to Black Box Testing.

### Regression Testing

Regression testing is done to ensure that any modifications in the system, its components, or a group of integrated units work efficiently, does not damage other related modules and does not produce unexpected results. It also belongs to the Black box testing category.

## CHAPTER 9

**CONCLUSION**

The classification task is used on student database to predict the students division on the basis of previous database. As there are many approaches that are used for data classification, the Naïve Bayes Classifier and K Nearest Neighbour Classifier are used here. Information’s like Marks (grades), Attendance Percentage, Staff remarks (ratings), ragging, social media usage, extra-curricular activities participation , stress (peer pressure, family pressure, emotional stress) were collected from the student’s , to predict the performance at the end of the semester. This study will help to the students and the teachers to improve the division of the student. This study will also work to identify those students which needed special attention to reduce fail ration and taking appropriate action for the next semester examination. This can help the students improve in their academics, which eventually leads to a good performance in their end examinations. This could help in our country development by providing good and efficient engineers to the country.

## CHAPTER 10 FUTURE ENHANCEMENT

This work could be used by various institutes by increasing the real-life dataset and helping the students to perform better by identifying the factors which are hampering the growth of the student. This could improve the student’s performance as multiple factors (example stress factor, ragging) other than just the academic factors are considered. This could also help teachers to understand an individual and guide him/her in the right direction.

## CHAPTER 11 REFERENCES

1. Sharma,P.,Vaghela,D., and Parmar,K.,"Performance Prediction of Students using Distributed Data Mining,” IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems ICIIECS’15,2015.
2. Osmanbegovic,E., and Suljic,M.,"Data mining approach for prediciting student's performance," Economic Review – Journal of Economics and Business, Vol. X, Issue 1 ,2012.
3. Banumathi,A., and Pethalakshmi,A.,"A Novel Approach for Upgrading Indian Education by Using Data Mining Techniques," IEEE, DOI: [10.1109/ICTEE.2012.6208603](https://doi.org/10.1109/ICTEE.2012.6208603),2012.
4. Anoopkumar,M., and Rahman,A.,"A Review on Data Mining Techniques and Factor Used in Educational Data Mining to Predict Student Amelioration," IEEE, DOI: [10.1109/SAPIENCE.2016.7684113](https://doi.org/10.1109/SAPIENCE.2016.7684113),2016.

# APPENDIX

Students grade

{

"cells": [

{

"cell\_type": "code",

"execution\_count": 29,

"metadata": {},

"outputs": [],

"source": [

"import numpy as np\n",

"import pandas as pd\n",

"import sklearn\n",

"from sklearn import preprocessing"

]

},

{

"cell\_type": "code",

"execution\_count": 30,

"metadata": {},

"outputs": [],

"source": [

"train1 = pd.read\_csv('student-mat.csv')"

]

},

{

"cell\_type": "code",

"execution\_count": 31,

"metadata": {},

"outputs": [],

"source": [

"train2 = pd.read\_csv('student-por.csv')"

]

},

{

"cell\_type": "code",

"execution\_count": 32,

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>school</th>\n",

" <th>sex</th>\n",

" <th>age</th>\n",

" <th>address</th>\n",

" <th>famsize</th>\n",

" <th>Pstatus</th>\n",

" <th>Medu</th>\n",

" <th>Fedu</th>\n",

" <th>Mjob</th>\n",

" <th>Fjob</th>\n",

" <th>...</th>\n",

" <th>famrel</th>\n",

" <th>freetime</th>\n",

" <th>goout</th>\n",

" <th>Dalc</th>\n",

" <th>Walc</th>\n",

" <th>health</th>\n",

" <th>absences</th>\n",

" <th>G1</th>\n",

" <th>G2</th>\n",

" <th>G3</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>GP</td>\n",

" <td>F</td>\n",

" <td>18</td>\n",

" <td>U</td>\n",

" <td>GT3</td>\n",

" <td>A</td>\n",

" <td>4</td>\n",

" <td>4</td>\n",

" <td>at\_home</td>\n",

" <td>teacher</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>4</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>6</td>\n",

" <td>5</td>\n",

" <td>6</td>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>GP</td>\n",

" <td>F</td>\n",

" <td>17</td>\n",

" <td>U</td>\n",

" <td>GT3</td>\n",

" <td>T</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>at\_home</td>\n",

" <td>other</td>\n",

" <td>...</td>\n",

" <td>5</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>4</td>\n",

" <td>5</td>\n",

" <td>5</td>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>GP</td>\n",

" <td>F</td>\n",

" <td>15</td>\n",

" <td>U</td>\n",

" <td>LE3</td>\n",

" <td>T</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>at\_home</td>\n",

" <td>other</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>10</td>\n",

" <td>7</td>\n",

" <td>8</td>\n",

" <td>10</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>GP</td>\n",

" <td>F</td>\n",

" <td>15</td>\n",

" <td>U</td>\n",

" <td>GT3</td>\n",

" <td>T</td>\n",

" <td>4</td>\n",

" <td>2</td>\n",

" <td>health</td>\n",

" <td>services</td>\n",

" <td>...</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>5</td>\n",

" <td>2</td>\n",

" <td>15</td>\n",

" <td>14</td>\n",

" <td>15</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>GP</td>\n",

" <td>F</td>\n",

" <td>16</td>\n",

" <td>U</td>\n",

" <td>GT3</td>\n",

" <td>T</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>other</td>\n",

" <td>other</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>2</td>\n",

" <td>5</td>\n",

" <td>4</td>\n",

" <td>6</td>\n",

" <td>10</td>\n",

" <td>10</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 33 columns</p>\n",

"</div>"

],

"text/plain": [

" school sex age address famsize Pstatus Medu Fedu Mjob Fjob ... \\\n",

"0 GP F 18 U GT3 A 4 4 at\_home teacher ... \n",

"1 GP F 17 U GT3 T 1 1 at\_home other ... \n",

"2 GP F 15 U LE3 T 1 1 at\_home other ... \n",

"3 GP F 15 U GT3 T 4 2 health services ... \n",

"4 GP F 16 U GT3 T 3 3 other other ... \n",

"\n",

" famrel freetime goout Dalc Walc health absences G1 G2 G3 \n",

"0 4 3 4 1 1 3 6 5 6 6 \n",

"1 5 3 3 1 1 3 4 5 5 6 \n",

"2 4 3 2 2 3 3 10 7 8 10 \n",

"3 3 2 2 1 1 5 2 15 14 15 \n",

"4 4 3 2 1 2 5 4 6 10 10 \n",

"\n",

"[5 rows x 33 columns]"

]

},

"execution\_count": 32,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"train1.head()"

]

},

{

"cell\_type": "code",

"execution\_count": 33,

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>school</th>\n",

" <th>sex</th>\n",

" <th>age</th>\n",

" <th>address</th>\n",

" <th>famsize</th>\n",

" <th>Pstatus</th>\n",

" <th>Medu</th>\n",

" <th>Fedu</th>\n",

" <th>Mjob</th>\n",

" <th>Fjob</th>\n",

" <th>...</th>\n",

" <th>famrel</th>\n",

" <th>freetime</th>\n",

" <th>goout</th>\n",

" <th>Dalc</th>\n",

" <th>Walc</th>\n",

" <th>health</th>\n",

" <th>absences</th>\n",

" <th>G1</th>\n",

" <th>G2</th>\n",

" <th>G3</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>GP</td>\n",

" <td>F</td>\n",

" <td>18</td>\n",

" <td>U</td>\n",

" <td>GT3</td>\n",

" <td>A</td>\n",

" <td>4</td>\n",

" <td>4</td>\n",

" <td>at\_home</td>\n",

" <td>teacher</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>4</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>4</td>\n",

" <td>0</td>\n",

" <td>11</td>\n",

" <td>11</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>GP</td>\n",

" <td>F</td>\n",

" <td>17</td>\n",

" <td>U</td>\n",

" <td>GT3</td>\n",

" <td>T</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>at\_home</td>\n",

" <td>other</td>\n",

" <td>...</td>\n",

" <td>5</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>9</td>\n",

" <td>11</td>\n",

" <td>11</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>GP</td>\n",

" <td>F</td>\n",

" <td>15</td>\n",

" <td>U</td>\n",

" <td>LE3</td>\n",

" <td>T</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>at\_home</td>\n",

" <td>other</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>6</td>\n",

" <td>12</td>\n",

" <td>13</td>\n",

" <td>12</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>GP</td>\n",

" <td>F</td>\n",

" <td>15</td>\n",

" <td>U</td>\n",

" <td>GT3</td>\n",

" <td>T</td>\n",

" <td>4</td>\n",

" <td>2</td>\n",

" <td>health</td>\n",

" <td>services</td>\n",

" <td>...</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>5</td>\n",

" <td>0</td>\n",

" <td>14</td>\n",

" <td>14</td>\n",

" <td>14</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>GP</td>\n",

" <td>F</td>\n",

" <td>16</td>\n",

" <td>U</td>\n",

" <td>GT3</td>\n",

" <td>T</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>other</td>\n",

" <td>other</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>2</td>\n",

" <td>5</td>\n",

" <td>0</td>\n",

" <td>11</td>\n",

" <td>13</td>\n",

" <td>13</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 33 columns</p>\n",

"</div>"

],

"text/plain": [

" school sex age address famsize Pstatus Medu Fedu Mjob Fjob ... \\\n",

"0 GP F 18 U GT3 A 4 4 at\_home teacher ... \n",

"1 GP F 17 U GT3 T 1 1 at\_home other ... \n",

"2 GP F 15 U LE3 T 1 1 at\_home other ... \n",

"3 GP F 15 U GT3 T 4 2 health services ... \n",

"4 GP F 16 U GT3 T 3 3 other other ... \n",

"\n",

" famrel freetime goout Dalc Walc health absences G1 G2 G3 \n",

"0 4 3 4 1 1 3 4 0 11 11 \n",

"1 5 3 3 1 1 3 2 9 11 11 \n",

"2 4 3 2 2 3 3 6 12 13 12 \n",

"3 3 2 2 1 1 5 0 14 14 14 \n",

"4 4 3 2 1 2 5 0 11 13 13 \n",

"\n",

"[5 rows x 33 columns]"

]

},

"execution\_count": 33,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"train2.head()"

]

},

{

"cell\_type": "code",

"execution\_count": 34,

"metadata": {},

"outputs": [],

"source": [

"# train1.info()\n",

"# train2.info()"

]

},

{

"cell\_type": "code",

"execution\_count": 35,

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"(395, 33)\n",

"(649, 33)\n"

]

}

],

"source": [

"print(train1.shape)\n",

"print(train2.shape)"

]

},

{

"cell\_type": "code",

"execution\_count": 36,

"metadata": {},

"outputs": [],

"source": [

"# train1.isnull().sum()"

]

},

{

"cell\_type": "code",

"execution\_count": 37,

"metadata": {},

"outputs": [],

"source": [

"# train2.isnull().sum()"

]

},

{

"cell\_type": "code",

"execution\_count": 72,

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"LabelEncoder()"

]

},

"execution\_count": 72,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"le=preprocessing.LabelEncoder()\n",

"le.fit(train1['school'])"

]

},

{

"cell\_type": "code",

"execution\_count": 73,

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"array([0, 1])"

]

},

"execution\_count": 73,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"le.classes\_"

]

},

{

"cell\_type": "code",

"execution\_count": 74,

"metadata": {},

"outputs": [],

"source": [

"train1['school']=le.transform(train1['school'])"

]

},

{

"cell\_type": "code",

"execution\_count": 75,

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>school</th>\n",

" <th>sex</th>\n",

" <th>age</th>\n",

" <th>address</th>\n",

" <th>famsize</th>\n",

" <th>Pstatus</th>\n",

" <th>Medu</th>\n",

" <th>Fedu</th>\n",

" <th>Mjob</th>\n",

" <th>Fjob</th>\n",

" <th>...</th>\n",

" <th>famrel</th>\n",

" <th>freetime</th>\n",

" <th>goout</th>\n",

" <th>Dalc</th>\n",

" <th>Walc</th>\n",

" <th>health</th>\n",

" <th>absences</th>\n",

" <th>G1</th>\n",

" <th>G2</th>\n",

" <th>G3</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>18</td>\n",

" <td>U</td>\n",

" <td>GT3</td>\n",

" <td>A</td>\n",

" <td>4</td>\n",

" <td>4</td>\n",

" <td>at\_home</td>\n",

" <td>teacher</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>4</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>6</td>\n",

" <td>5</td>\n",

" <td>6</td>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>17</td>\n",

" <td>U</td>\n",

" <td>GT3</td>\n",

" <td>T</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>at\_home</td>\n",

" <td>other</td>\n",

" <td>...</td>\n",

" <td>5</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>4</td>\n",

" <td>5</td>\n",

" <td>5</td>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>15</td>\n",

" <td>U</td>\n",

" <td>LE3</td>\n",

" <td>T</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>at\_home</td>\n",

" <td>other</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>10</td>\n",

" <td>7</td>\n",

" <td>8</td>\n",

" <td>10</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>15</td>\n",

" <td>U</td>\n",

" <td>GT3</td>\n",

" <td>T</td>\n",

" <td>4</td>\n",

" <td>2</td>\n",

" <td>health</td>\n",

" <td>services</td>\n",

" <td>...</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>5</td>\n",

" <td>2</td>\n",

" <td>15</td>\n",

" <td>14</td>\n",

" <td>15</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>16</td>\n",

" <td>U</td>\n",

" <td>GT3</td>\n",

" <td>T</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>other</td>\n",

" <td>other</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>2</td>\n",

" <td>5</td>\n",

" <td>4</td>\n",

" <td>6</td>\n",

" <td>10</td>\n",

" <td>10</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 33 columns</p>\n",

"</div>"

],

"text/plain": [

" school sex age address famsize Pstatus Medu Fedu Mjob Fjob \\\n",

"0 0 0 18 U GT3 A 4 4 at\_home teacher \n",

"1 0 0 17 U GT3 T 1 1 at\_home other \n",

"2 0 0 15 U LE3 T 1 1 at\_home other \n",

"3 0 1 15 U GT3 T 4 2 health services \n",

"4 0 0 16 U GT3 T 3 3 other other \n",

"\n",

" ... famrel freetime goout Dalc Walc health absences G1 G2 G3 \n",

"0 ... 4 3 4 1 1 3 6 5 6 6 \n",

"1 ... 5 3 3 1 1 3 4 5 5 6 \n",

"2 ... 4 3 2 2 3 3 10 7 8 10 \n",

"3 ... 3 2 2 1 1 5 2 15 14 15 \n",

"4 ... 4 3 2 1 2 5 4 6 10 10 \n",

"\n",

"[5 rows x 33 columns]"

]

},

"execution\_count": 75,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"train1.head()"

]

},

{

"cell\_type": "code",

"execution\_count": 76,

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"Index(['school', 'sex', 'age', 'address', 'famsize', 'Pstatus', 'Medu', 'Fedu',\n",

" 'Mjob', 'Fjob', 'reason', 'guardian', 'traveltime', 'studytime',\n",

" 'failures', 'schoolsup', 'famsup', 'paid', 'activities', 'nursery',\n",

" 'higher', 'internet', 'romantic', 'famrel', 'freetime', 'goout', 'Dalc',\n",

" 'Walc', 'health', 'absences', 'G1', 'G2', 'G3'],\n",

" dtype='object')"

]

},

"execution\_count": 76,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"train1.columns\n",

"\n"

]

},

{

"cell\_type": "code",

"execution\_count": 77,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['sex'])\n",

"train1['sex']=le.transform(train1['sex'])"

]

},

{

"cell\_type": "code",

"execution\_count": 78,

"metadata": {},

"outputs": [],

"source": [

"# train1.head()"

]

},

{

"cell\_type": "code",

"execution\_count": 79,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['address'])\n",

"train1['address']=le.transform(train1['address'])"

]

},

{

"cell\_type": "code",

"execution\_count": 80,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['famsize'])\n",

"train1['famsize']=le.transform(train1['famsize'])"

]

},

{

"cell\_type": "code",

"execution\_count": 81,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['Pstatus'])\n",

"train1['Pstatus']=le.transform(train1['Pstatus'])"

]

},

{

"cell\_type": "code",

"execution\_count": 82,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['Mjob'])\n",

"train1['Mjob']=le.transform(train1['Mjob'])"

]

},

{

"cell\_type": "code",

"execution\_count": 83,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['Fjob'])\n",

"train1['Fjob']=le.transform(train1['Fjob'])"

]

},

{

"cell\_type": "code",

"execution\_count": 84,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['reason'])\n",

"train1['reason']=le.transform(train1['reason'])"

]

},

{

"cell\_type": "code",

"execution\_count": 85,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['guardian'])\n",

"train1['guardian']=le.transform(train1['guardian'])"

]

},

{

"cell\_type": "code",

"execution\_count": 86,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['schoolsup'])\n",

"train1['schoolsup']=le.transform(train1['schoolsup'])"

]

},

{

"cell\_type": "code",

"execution\_count": 87,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['famsup'])\n",

"train1['famsup']=le.transform(train1['famsup'])"

]

},

{

"cell\_type": "code",

"execution\_count": 88,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['paid'])\n",

"train1['paid']=le.transform(train1['paid'])"

]

},

{

"cell\_type": "code",

"execution\_count": 89,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['activities'])\n",

"train1['activities']=le.transform(train1['activities'])"

]

},

{

"cell\_type": "code",

"execution\_count": 90,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['nursery'])\n",

"train1['nursery']=le.transform(train1['nursery'])"

]

},

{

"cell\_type": "code",

"execution\_count": 91,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['higher'])\n",

"train1['higher']=le.transform(train1['higher'])"

]

},

{

"cell\_type": "code",

"execution\_count": 92,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['internet'])\n",

"train1['internet']=le.transform(train1['internet'])"

]

},

{

"cell\_type": "code",

"execution\_count": 93,

"metadata": {},

"outputs": [],

"source": [

"le.fit(train1['romantic'])\n",

"train1['romantic']=le.transform(train1['romantic'])"

]

},

{

"cell\_type": "code",

"execution\_count": 95,

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>school</th>\n",

" <th>sex</th>\n",

" <th>age</th>\n",

" <th>address</th>\n",

" <th>famsize</th>\n",

" <th>Pstatus</th>\n",

" <th>Medu</th>\n",

" <th>Fedu</th>\n",

" <th>Mjob</th>\n",

" <th>Fjob</th>\n",

" <th>...</th>\n",

" <th>famrel</th>\n",

" <th>freetime</th>\n",

" <th>goout</th>\n",

" <th>Dalc</th>\n",

" <th>Walc</th>\n",

" <th>health</th>\n",

" <th>absences</th>\n",

" <th>G1</th>\n",

" <th>G2</th>\n",

" <th>G3</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>18</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>4</td>\n",

" <td>4</td>\n",

" <td>0</td>\n",

" <td>4</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>4</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>6</td>\n",

" <td>5</td>\n",

" <td>6</td>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>17</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>2</td>\n",

" <td>...</td>\n",

" <td>5</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>4</td>\n",

" <td>5</td>\n",

" <td>5</td>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>15</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>2</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>10</td>\n",

" <td>7</td>\n",

" <td>8</td>\n",

" <td>10</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>15</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>4</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>...</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>5</td>\n",

" <td>2</td>\n",

" <td>15</td>\n",

" <td>14</td>\n",

" <td>15</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>16</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>2</td>\n",

" <td>5</td>\n",

" <td>4</td>\n",

" <td>6</td>\n",

" <td>10</td>\n",

" <td>10</td>\n",

" </tr>\n",

" <tr>\n",

" <th>5</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>16</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>...</td>\n",

" <td>5</td>\n",

" <td>4</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>2</td>\n",

" <td>5</td>\n",

" <td>10</td>\n",

" <td>15</td>\n",

" <td>15</td>\n",

" <td>15</td>\n",

" </tr>\n",

" <tr>\n",

" <th>6</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>16</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>4</td>\n",

" <td>4</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>0</td>\n",

" <td>12</td>\n",

" <td>12</td>\n",

" <td>11</td>\n",

" </tr>\n",

" <tr>\n",

" <th>7</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>17</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>4</td>\n",

" <td>4</td>\n",

" <td>2</td>\n",

" <td>4</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>1</td>\n",

" <td>4</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>6</td>\n",

" <td>6</td>\n",

" <td>5</td>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>8</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>15</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>...</td>\n",

" <td>4</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>16</td>\n",

" <td>18</td>\n",

" <td>19</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>15</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>4</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>...</td>\n",

" <td>5</td>\n",

" <td>5</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>5</td>\n",

" <td>0</td>\n",

" <td>14</td>\n",

" <td>15</td>\n",

" <td>15</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>10 rows × 33 columns</p>\n",

"</div>"

],

"text/plain": [

" school sex age address famsize Pstatus Medu Fedu Mjob Fjob ... \\\n",

"0 0 0 18 1 0 0 4 4 0 4 ... \n",

"1 0 0 17 1 0 1 1 1 0 2 ... \n",

"2 0 0 15 1 1 1 1 1 0 2 ... \n",

"3 0 1 15 1 0 1 4 2 1 3 ... \n",

"4 0 0 16 1 0 1 3 3 2 2 ... \n",

"5 0 0 16 1 1 1 4 3 3 2 ... \n",

"6 0 0 16 1 1 1 2 2 2 2 ... \n",

"7 0 0 17 1 0 0 4 4 2 4 ... \n",

"8 0 0 15 1 1 0 3 2 3 2 ... \n",

"9 0 0 15 1 0 1 3 4 2 2 ... \n",

"\n",

" famrel freetime goout Dalc Walc health absences G1 G2 G3 \n",

"0 4 3 4 1 1 3 6 5 6 6 \n",

"1 5 3 3 1 1 3 4 5 5 6 \n",

"2 4 3 2 2 3 3 10 7 8 10 \n",

"3 3 2 2 1 1 5 2 15 14 15 \n",

"4 4 3 2 1 2 5 4 6 10 10 \n",

"5 5 4 2 1 2 5 10 15 15 15 \n",

"6 4 4 4 1 1 3 0 12 12 11 \n",

"7 4 1 4 1 1 1 6 6 5 6 \n",

"8 4 2 2 1 1 1 0 16 18 19 \n",

"9 5 5 1 1 1 5 0 14 15 15 \n",

"\n",

"[10 rows x 33 columns]"

]

},

"execution\_count": 95,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"train1.head(10)"

]

},

{

"cell\_type": "code",

"execution\_count": 96,

"metadata": {},

"outputs": [],

"source": [

"train1\_y=train1['G3']"

]

},

{

"cell\_type": "code",

"execution\_count": 100,

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>school</th>\n",

" <th>sex</th>\n",

" <th>age</th>\n",

" <th>address</th>\n",

" <th>famsize</th>\n",

" <th>Pstatus</th>\n",

" <th>Medu</th>\n",

" <th>Fedu</th>\n",

" <th>Mjob</th>\n",

" <th>Fjob</th>\n",

" <th>...</th>\n",

" <th>romantic</th>\n",

" <th>famrel</th>\n",

" <th>freetime</th>\n",

" <th>goout</th>\n",

" <th>Dalc</th>\n",

" <th>Walc</th>\n",

" <th>health</th>\n",

" <th>absences</th>\n",

" <th>G1</th>\n",

" <th>G2</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>18</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>4</td>\n",

" <td>4</td>\n",

" <td>0</td>\n",

" <td>4</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>4</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>6</td>\n",

" <td>5</td>\n",

" <td>6</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>17</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>2</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>5</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>4</td>\n",

" <td>5</td>\n",

" <td>5</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>15</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>2</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>10</td>\n",

" <td>7</td>\n",

" <td>8</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>15</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>4</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>...</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>5</td>\n",

" <td>2</td>\n",

" <td>15</td>\n",

" <td>14</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>16</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>3</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>2</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>4</td>\n",

" <td>3</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>2</td>\n",

" <td>5</td>\n",

" <td>4</td>\n",

" <td>6</td>\n",

" <td>10</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 32 columns</p>\n",

"</div>"

],

"text/plain": [

" school sex age address famsize Pstatus Medu Fedu Mjob Fjob ... \\\n",

"0 0 0 18 1 0 0 4 4 0 4 ... \n",

"1 0 0 17 1 0 1 1 1 0 2 ... \n",

"2 0 0 15 1 1 1 1 1 0 2 ... \n",

"3 0 1 15 1 0 1 4 2 1 3 ... \n",

"4 0 0 16 1 0 1 3 3 2 2 ... \n",

"\n",

" romantic famrel freetime goout Dalc Walc health absences G1 G2 \n",

"0 0 4 3 4 1 1 3 6 5 6 \n",

"1 0 5 3 3 1 1 3 4 5 5 \n",

"2 0 4 3 2 2 3 3 10 7 8 \n",

"3 1 3 2 2 1 1 5 2 15 14 \n",

"4 0 4 3 2 1 2 5 4 6 10 \n",

"\n",

"[5 rows x 32 columns]"

]

},

"execution\_count": 100,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"train1 = train1.drop(labels=['G3'],axis=1)\n",

"train1.head()"

]

},

{

"cell\_type": "code",

"execution\_count": 101,

"metadata": {},

"outputs": [

{

"ename": "NameError",

"evalue": "name 'LogisticRegression' is not defined",

"output\_type": "error",

"traceback": [

"\u001b[0;31m---------------------------------------------------------------------------\u001b[0m",

"\u001b[0;31mNameError\u001b[0m Traceback (most recent call last)",

"\u001b[0;32m<ipython-input-101-7bf3edf56497>\u001b[0m in \u001b[0;36m<module>\u001b[0;34m\u001b[0m\n\u001b[0;32m----> 1\u001b[0;31m \u001b[0mlg\u001b[0m\u001b[0;34m=\u001b[0m\u001b[0mLogisticRegression\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m",

"\u001b[0;31mNameError\u001b[0m: name 'LogisticRegression' is not defined"

]

}

],

"source": [

"from sklearn.linear\_model import LogisticRegression\n",

"lr = LogisticRegression()\n"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {},

"outputs": [],

"source": []

}

],

"metadata": {

"kernelspec": {

"display\_name": "Python 3",

"language": "python",

"name": "python3"

},

"language\_info": {

"codemirror\_mode": {

"name": "ipython",

"version": 3

},

"file\_extension": ".py",

"mimetype": "text/x-python",

"name": "python",

"nbconvert\_exporter": "python",

"pygments\_lexer": "ipython3",

"version": "3.6.6"

}

},

"nbformat": 4,

"nbformat\_minor": 2

}