A

Major Project

On

STUDENT'S PERFORMANCE ANALYSIS SYSTEM

Submitted To

Jawaharlal Nehru Technological University, Hyderabad

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

VINAY DEVABHAKTUNI	**187R1A05J3**
K. SHARATH REDDY	**187R1A05K9**
V. SHIVA TEJA	**187R1A05P6**

Under the esteemed guidance of

Mrs. G. KAVITHA REDDY

(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi) Recognized Under Section 2(f) & 12(B) of the UGC Act.1956,

Kandlakoya (V), Medchal Road, Hyderabad-501401.

2018-2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project entitled "**STUDENTS PERFORMANCE ANALYSIS SYSTEM**" being submitted by **VINAY DEVABHAKTUMI (187R1A05J3),K. SHARATH REDDY (187R1A05K9), V. SHIVA TEJA (187R1A05P6)** in partial fulfillment of the requirements for the award of the degree of B. Tech in Computer Science and Engineering of the Jawaharlal Nehru Technological University Hyderabad, is a record of Bonafide work carried out by him/her under our guidance and supervision during the year 2021-22.

The result embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Mrs.G. KAVITHA REDDY **DR. A. Raji Reddy**

Assistant professor **Director**

Internal Guide

Dr. K. Srujan Raju **EXTERNAL EXAMINER**
HOD

Submitted on viva voice Examination held on_____

ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project. We take this opportunity to express my profound gratitude and deep regard to my guide

Mrs.G.KAVITHA , Assistant. Professor for her exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to Project Review Committee (PRC) Coordinators: **Mr. A. Uday Kiran, Mr. J. NarasimhaRao, Dr. T. S. Mastan Rao, Mrs. G. Latha ,Mr. A. Kiran Kumar** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to the Head of the Department **Dr. K. Srujan Raju** for providing excellent infrastructure and a nice atmosphere for completing this project successfully.

We are obliged to our Director **Dr. A. Raji Reddy** for being cooperative throughout the course of this project. We would like to express our sincere gratitude to our Chairman Sri. **Ch. Gopal Reddy** for his encouragement throughout the course of this project

The guidance and support received from all the members of **CMR TECHNICAL CAMPUS** who contributed and who are contributing to this project, was vital for the success of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement without which this assignment would not be possible. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project.

VINAY DEVABHKTUNI (187R1A05J3)

K. SHARATH REDYY (187R1A05K9)

V. SHIVA TEJA (187R1A05P6)

ABSTRACT

Towards automation to do mundane tasks and the expectations for students already equipped with good programming skills is on the rise. In parallel, there has been a rising number of students who find it difficult to attain the skills necessary in order to get the dream IT job they desire. The aim of this project is to bridge the gap between the employer and the future employee of the company by the use of SPAS at college level. Student performance analysis system (SPAS) is an online web application system which enables students to know prior hand if their level of skills for the placement is enough to get placed or not, given the necessary inputs. SPAS has an intelligent learning algorithm which utilises a rich database, analyses the records of previous students' traits and develops a model for further prediction. The performance evaluation of students by SPAS is by the cumulative predictor algorithm involving generation of several random forest trees on the available data. SPAS learns and creates its model reaching higher accuracy with increasing data availability.

LIST OF FIGURES

LIST OF SCREENSHOTS

TABLE OF CONTENT

1. INTRODUCTION

1. INTRODUCTION

1.1 PROJECT SCOPE

Educational data mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. There are several data regarding the students which stay unused with untapped potential of data mining which could revolution is the field of education. Since the ultimate aim of an educational institution is to create a pool of skilled professionals to take on the society to a next upgraded level, they need to create an environment for their students to grow in every vertical by giving them right exposure and training. Most of the educational institutions, maintain huge databases of students and the information keeps on increasing with time, but there is no action taken to gain knowledge from it. DM has the suitable techniques in mining the data to discover new information and knowledge about students. DM provides various methods for analysis which include classification, clustering, and association rules.

1.2 PROJECT PURPOSE

The student academic performance is usually stored in student management system, in different formats such as files, document, records, images and other formats. These available students' data could be extracted to produce useful information. However, the increasing amount of students' data becomes hard to be analysed by using traditional statistic techniques and database management tools . Thus, a tool is necessary for universities to extract the useful information. This useful information could be used to predict the students' performance

1.3 PROJECT FEATURES

i. To develop a system for students' performance analysis.

ii. To assist the lecturers in analysing and predicting student performance in course"System Analysis and Design" by using data mining technique in the proposed system.

iii. To identify the factors that affect the students'performance in course "System Analysis and Design"

iv. To assist lecturers in keeping track of the students' progress throughout the semester

2.SYSTEM ANALYSIS

2.SYSTEM ANALYSIS

2.1 EXISTING SYSTEM:

The data were collected from eight year period intakes from July 2006/2007 until July 2013/2014 that contains the students' demographics, previous academic records, and family background information. DT, NB, and RB classification techniques are applied to the students' data in order to produce the best SAP prediction model. The experiment result shows the RB is a best model among the other techniques by receiving the highest accuracy value of 71.3%. The extracted knowledge from prediction model will be used to identify and profile the student to determine the students' level of success in the first semester. This project acts as the basis of SPAS and gives a clear idea regarding the parameters involved in predicting students' performance.

2.1.1 DISADVANTAGES OF EXISTING SYSTEM:

1. Less performance
2. Prediction problems

2.2 Proposed System:

The aim of this project is to bridge the gap between the employer and the future employee of the company by the use of SPAS at college level. Student performance analysis system (SPAS) is an online web application system which enables students to know prior hand if their level of skills for the placement is enough to get placed or not, given the necessary inputs. SPAS has an intelligent learning algorithm which utilises a rich database, analyses the records of previous students' traits and develops a model for further prediction.

Advantages of proposed system:

1. SPAS learns and creates its model reaching higher accuracy with increasing data availability.

2.3 SYSTEM REQUIREMENTS:

SOFTWARE REQUIREMENTS

The functional requirements or the overall description documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraints and user documentation.

The appropriation of requirements and implementation constraints gives the general overview of the project in regards to what the areas of strength and deficit are and how to tackle them.

- **Python ide 3.7 version (or)**
- **Anaconda 3.7 (or)**
- **Jupiter (or)**
- **Google colab**

HARDWARE REQUIREMENTS

Minimum hardware requirements are very dependent on the particular software being developed by a given Enthought Python / Canopy / VS Code user. Applications that need to store large arrays/objects in memory will require more RAM, whereas applications that need to perform numerous calculations or tasks more quickly will require a faster processor.

- **Operating system : windows, linux**
- **Processor : minimum intel i3**
- **Ram : minimum 4 gb**
- **Hard disk : minimum 250gb**

2.4 FUNCTIONAL REQUIREMENTS

1.Data Collection

2.Data Preprocessing

3.Training And Testing

4.Modiling

5.Predicting

2.5 NON FUNCTIONAL REQUIREMENTS

NON-FUNCTIONAL REQUIREMENT (NFR) specifies the quality attribute of a software system. They judge the software system based on Responsiveness, Usability, Security, Portability and other non-functional standards that are critical to the success of the software system. Example of nonfunctional requirement, *"how fast does the website load?"* Failing to meet non-functional requirements can result in systems that fail to satisfy user needs. Non- functional Requirements allows you to impose constraints or restrictions on the design of the system across the various agile backlogs. Example, the site should load in 3 seconds when the number of simultaneous users are > 10000. Description of non-functional requirements is just as critical as a functional requirement.

- Usability requirement
- Serviceability requirement
- Manageability requirement
- Recoverability requirement
- Security requirement
- Data Integrity requirement
- Capacity requirement
- Availability requirement
- Scalability requirement
- Interoperability requirement
- Reliability requirement
- Maintainability requirement
- Regulatory requirement
- Environmental requirement

2.6 SYSTEM STUDY

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

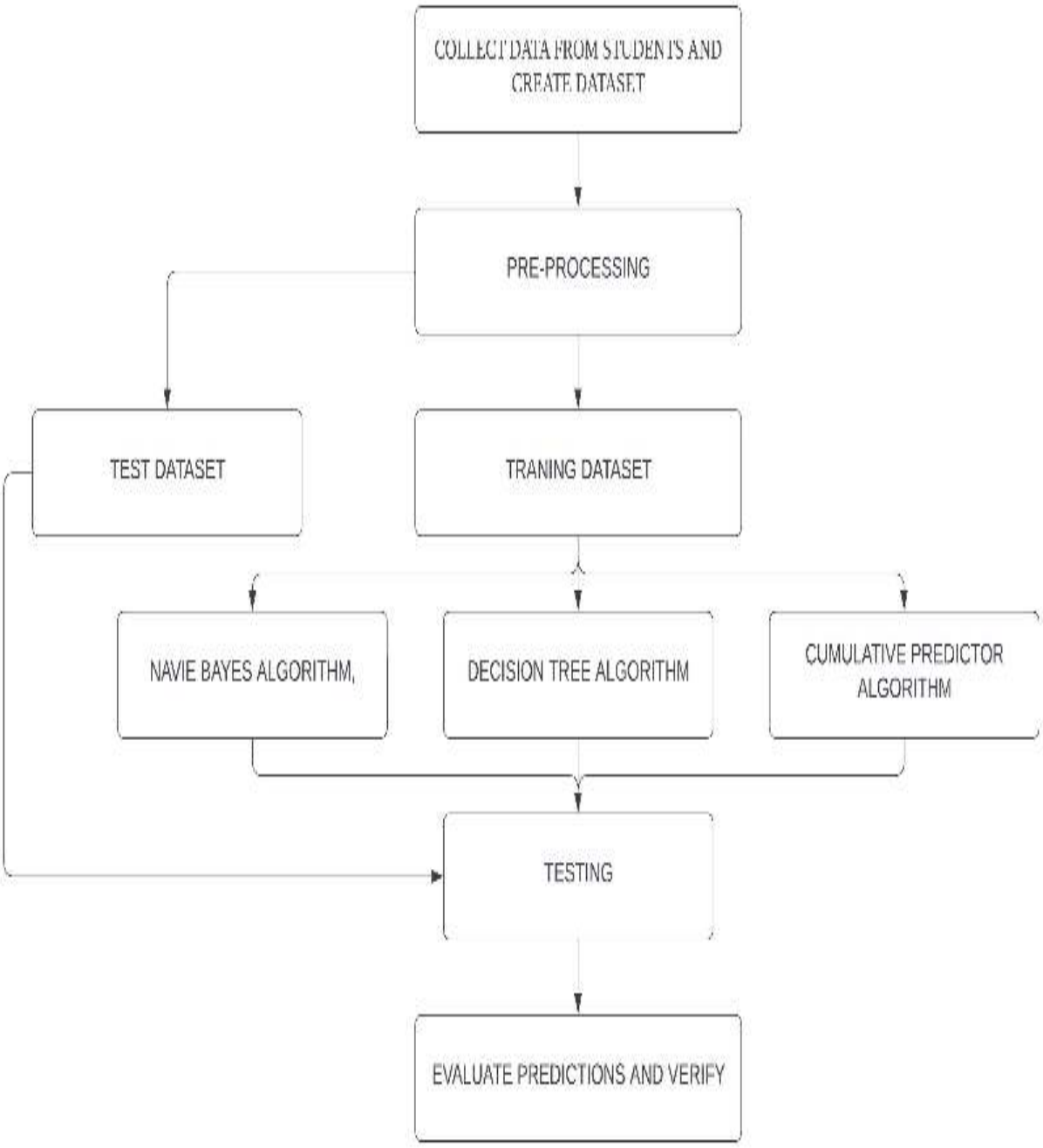
SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

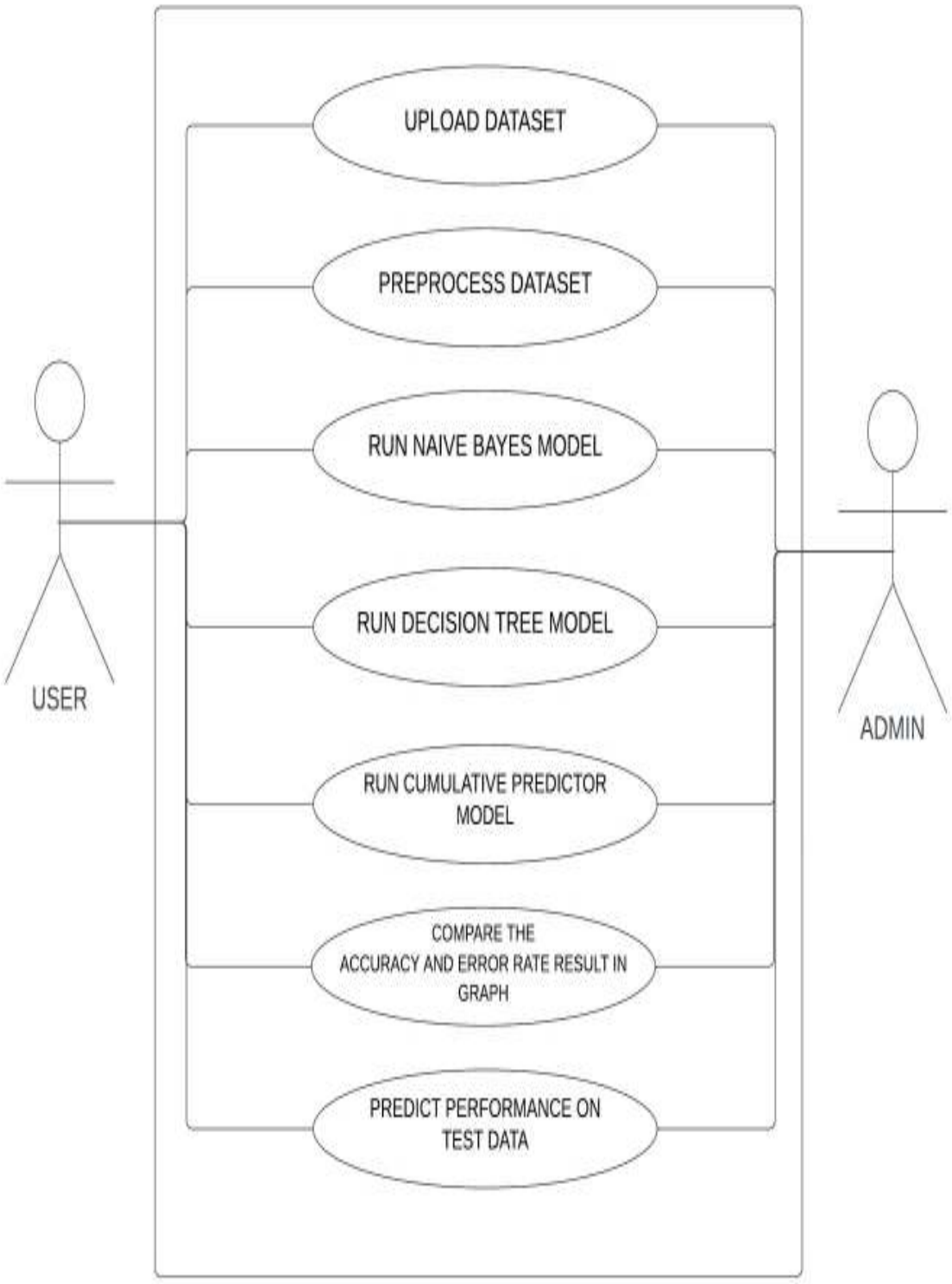
3.ARCHITECTURE

3 ARCHITECTURE

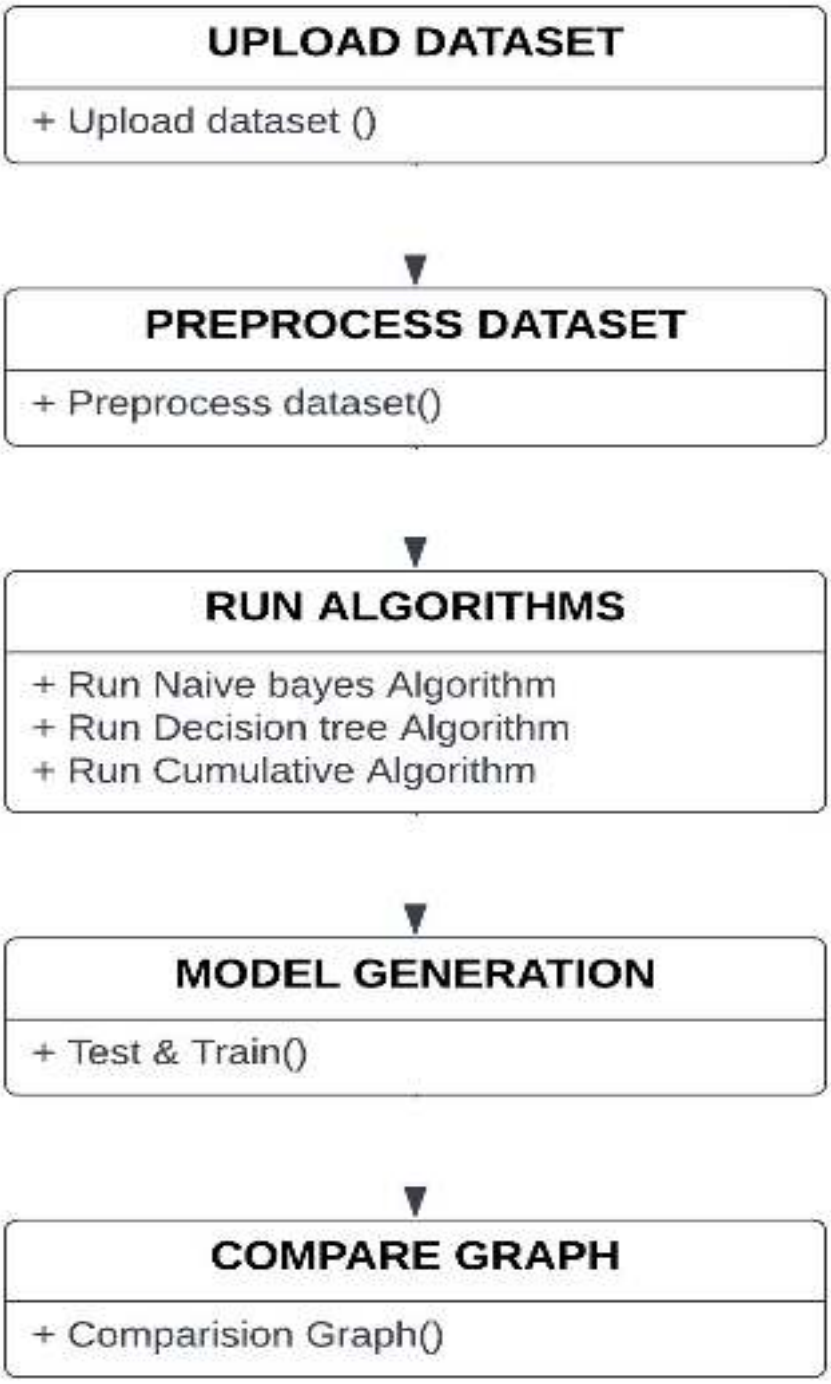
3.1 SYSTEM ARCHITECTURE



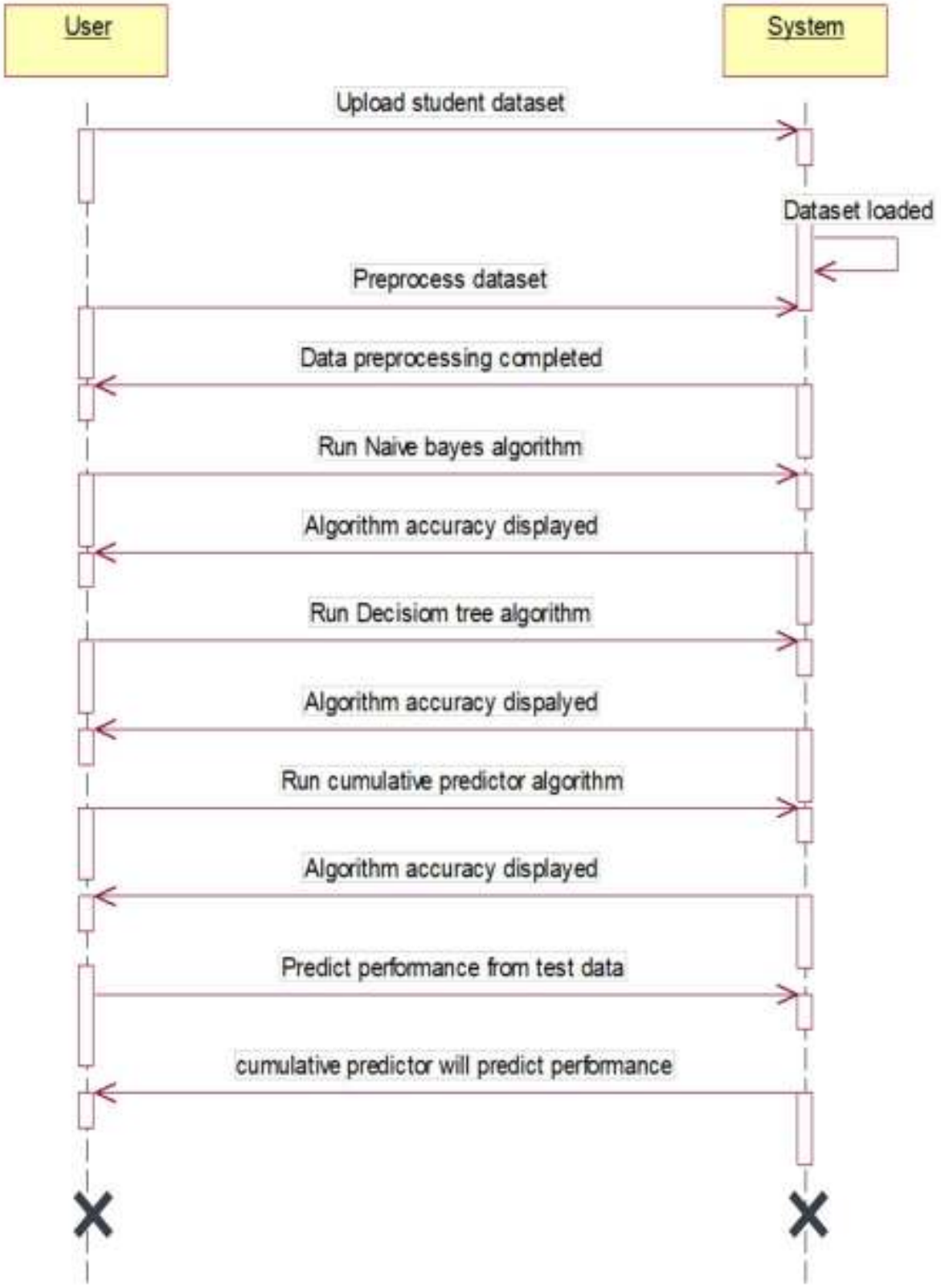
3.2 USE CASE DIAGRAM



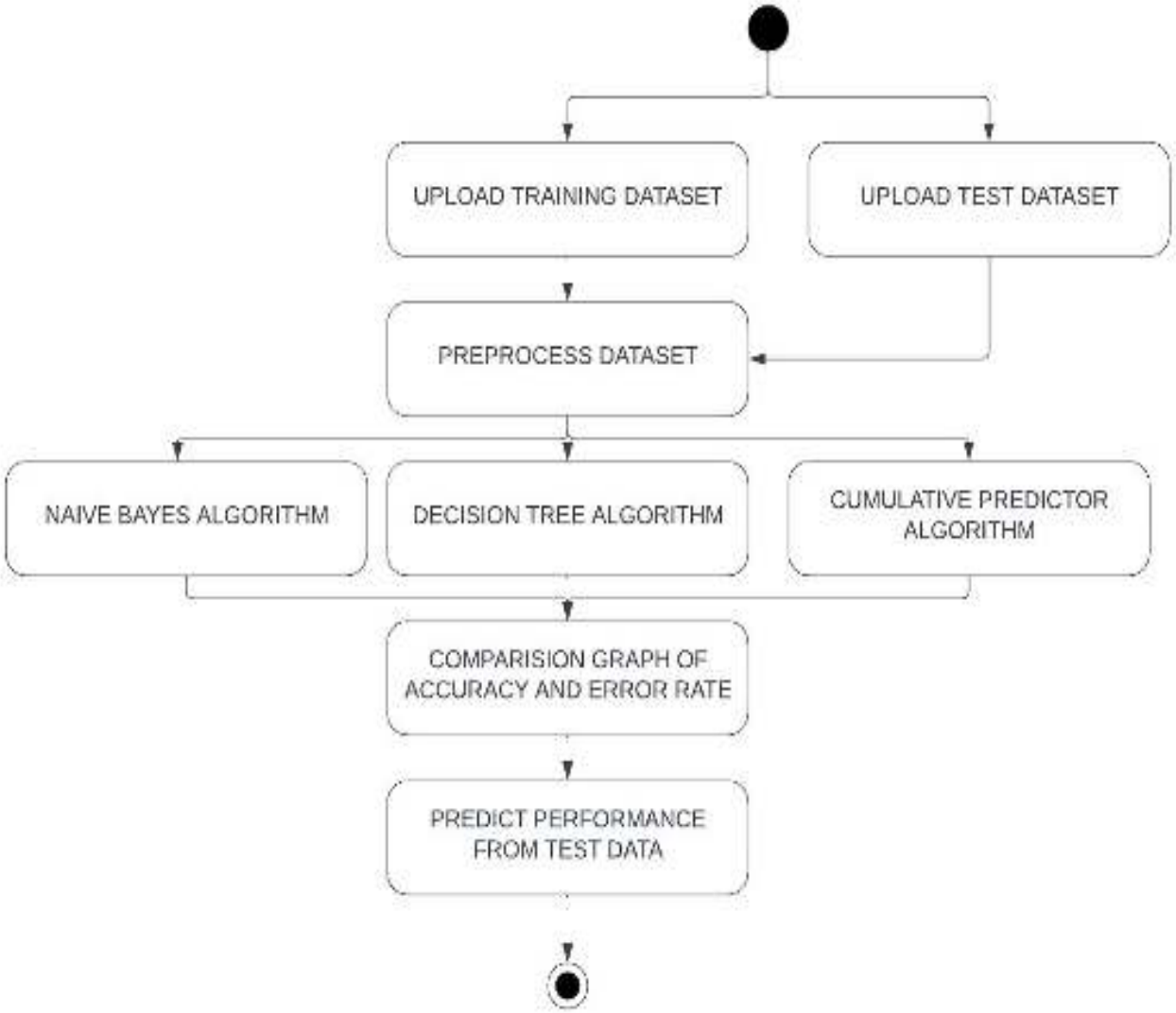
3.3 CLASS DIAGRAM



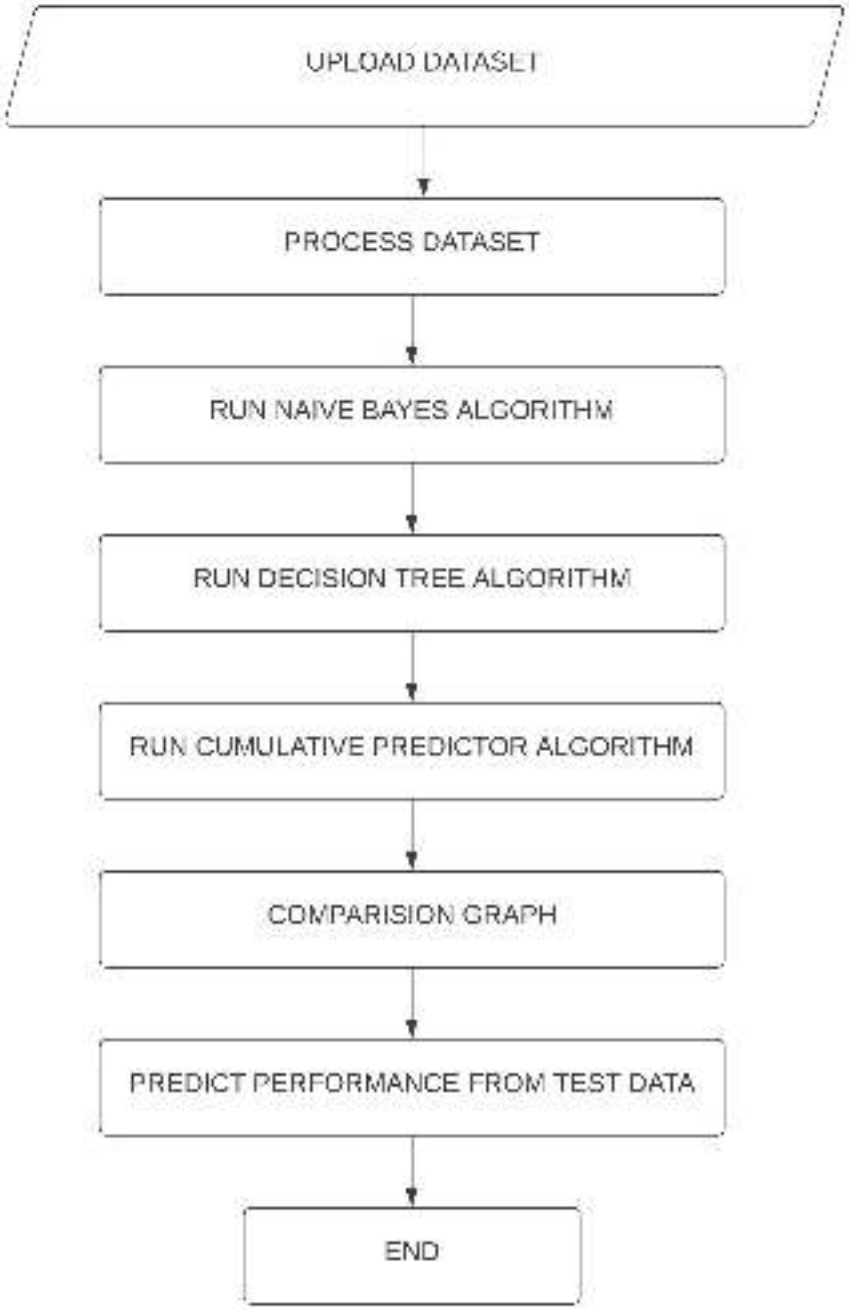
3.4 SEQUENCE DIAGRAM



3.5 ACTIVITYDIAGRAM



3.6 DATA FLOW DIAGRAM:



4.IMPLEMENTATION

4.1 SAMPLE CODE

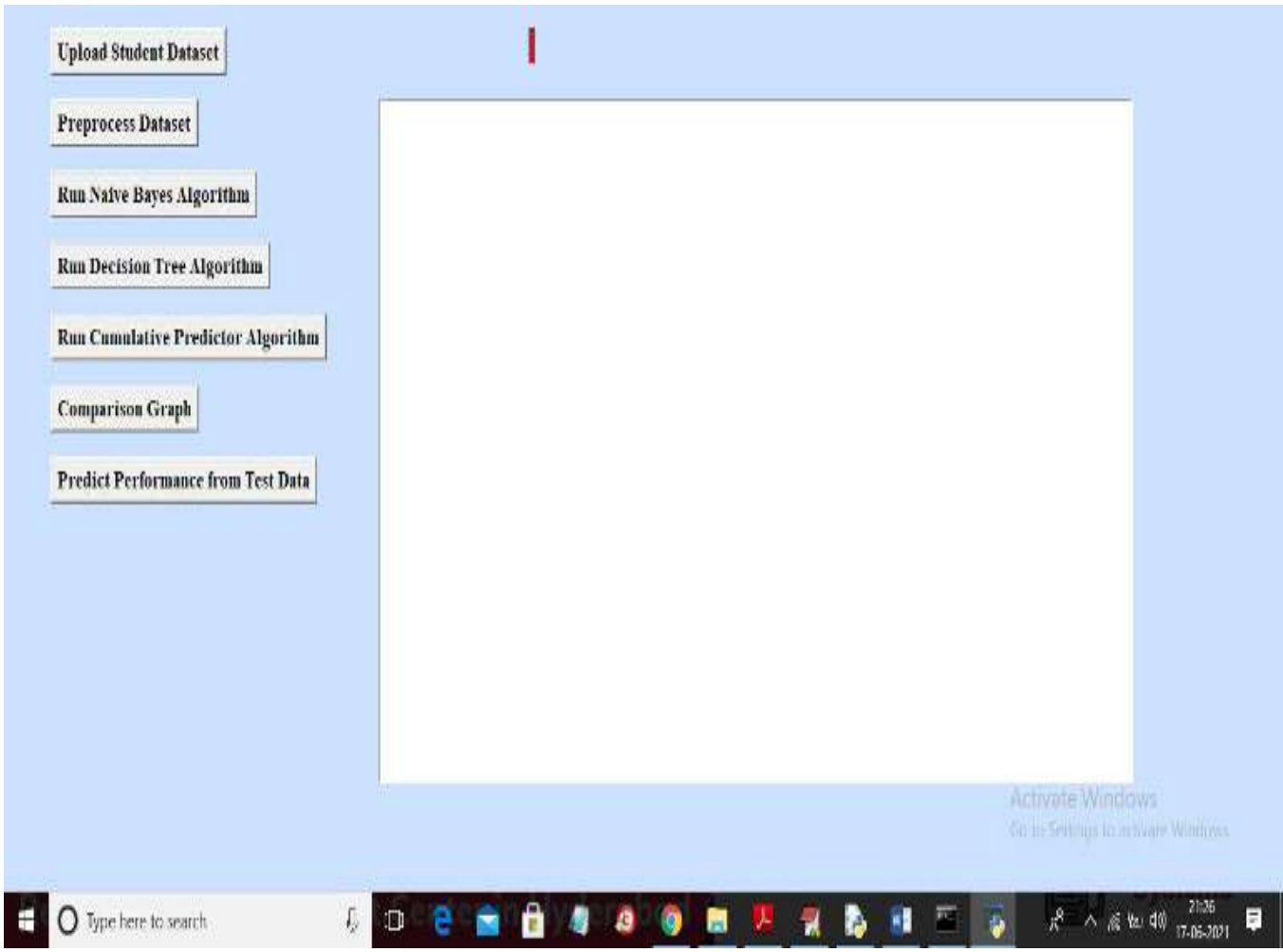
```
import
numpy
as np
        import pandas as pd

        """ Read data file as DataFrame """
        df = pd.read_csv("./data/student-mat.csv", sep=";")

        """ Import ML helpers """
        from sklearn.preprocessing import LabelEncoder
        from sklearn.model_selection import train_test_split
        from sklearn.metrics import confusion_matrix
        from sklearn.model_selection import GridSearchCV, cross_val_score

        from sklearn.pipeline import Pipeline
        from sklearn.feature_selection import SelectKBest, chi2
        from sklearn.svm import LinearSVC # Support Vector Machine Classifier
        model

        """ Split Data into Training and Testing Sets """
        def split_data(X, Y):
            return train_test_split(X, Y, test_size=0.2, random_state=17)

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

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

        """ False Fail Rate """
        def ffr(confusion_matrix):
            ff = confusion_matrix[1][0]
```

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

    return rate

""" Train Model and Print Score """
def train_and_score(X, y):
    X_train, X_test, y_train, y_test = split_data(X, y)

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

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

    clf.fit(X_train, y_train)

    confuse(y_test, clf.predict(X_test))
    print()

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

    # For each feature, encode to categorical values
    class_le = LabelEncoder()
    for column in df[["school", "sex", "address", "famsize", "Pstatus",
"Mjob", "Fjob", "reason", "guardian", "schoolsup", "famsup", "paid",
"activities", "nursery", "higher", "internet", "romantic"]].columns:
        df[column] = class_le.fit_transform(df[column].values)

    # Encode G1, G2, G3 as pass or fail binary values
    for i, row in df.iterrows():
        if row["G1"] >= 10:
            df["G1"][i] = 1
        else:
            df["G1"][i] = 0

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

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

    # Target values are G3
    y = df.pop("G3")

    # Feature set is remaining features
    X = df

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

    # Remove grade report 2
    X.drop(["G2"], axis = 1, inplace=True)
    print("\n\nModel Accuracy Knowing Only G1 Score")
    print("=====================================")
    train_and_score(X, y)

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



main()
```

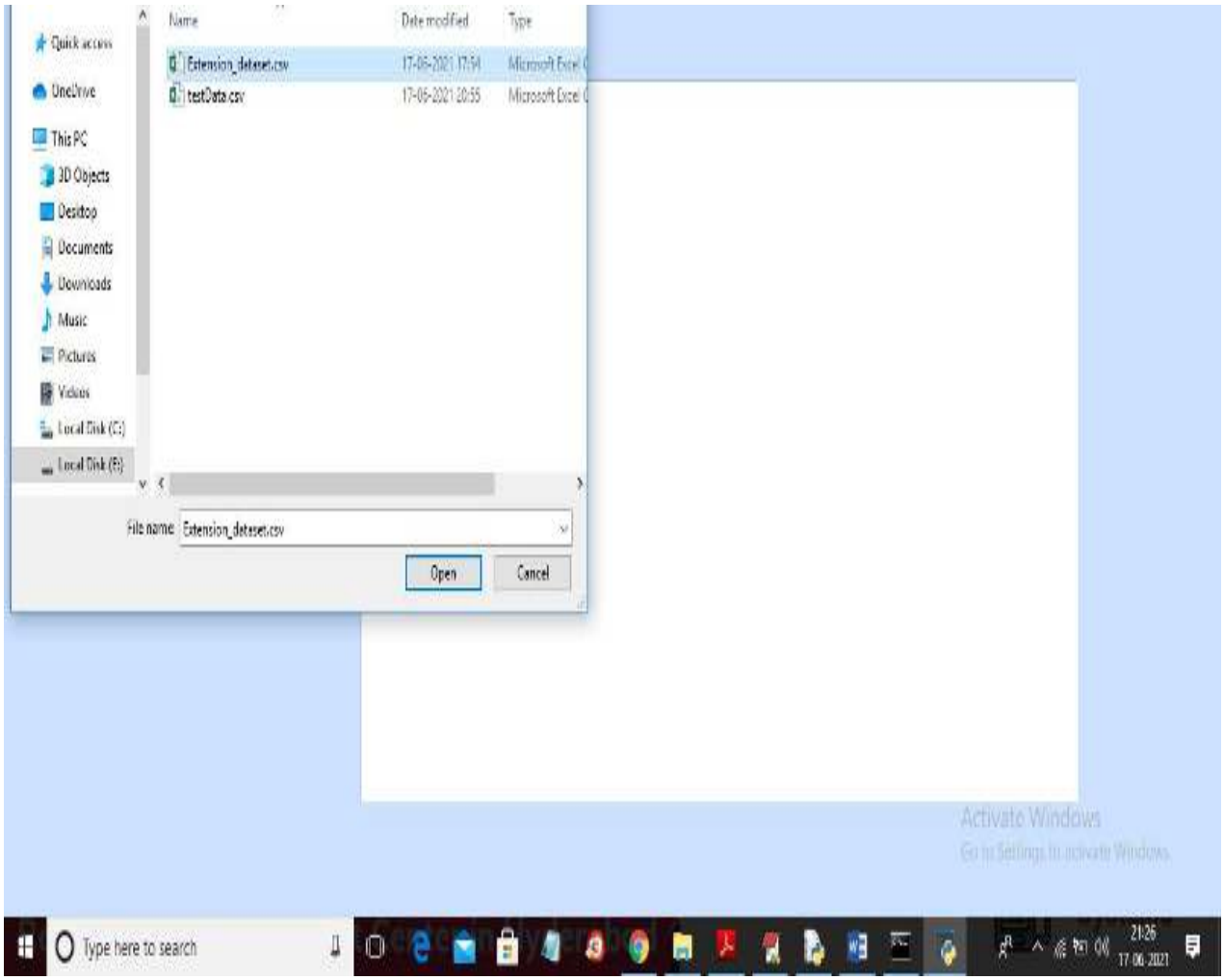
5.SCREENSHOTS

5.1 MAIN SCREEN AFTER STARTING THE PROGRAM

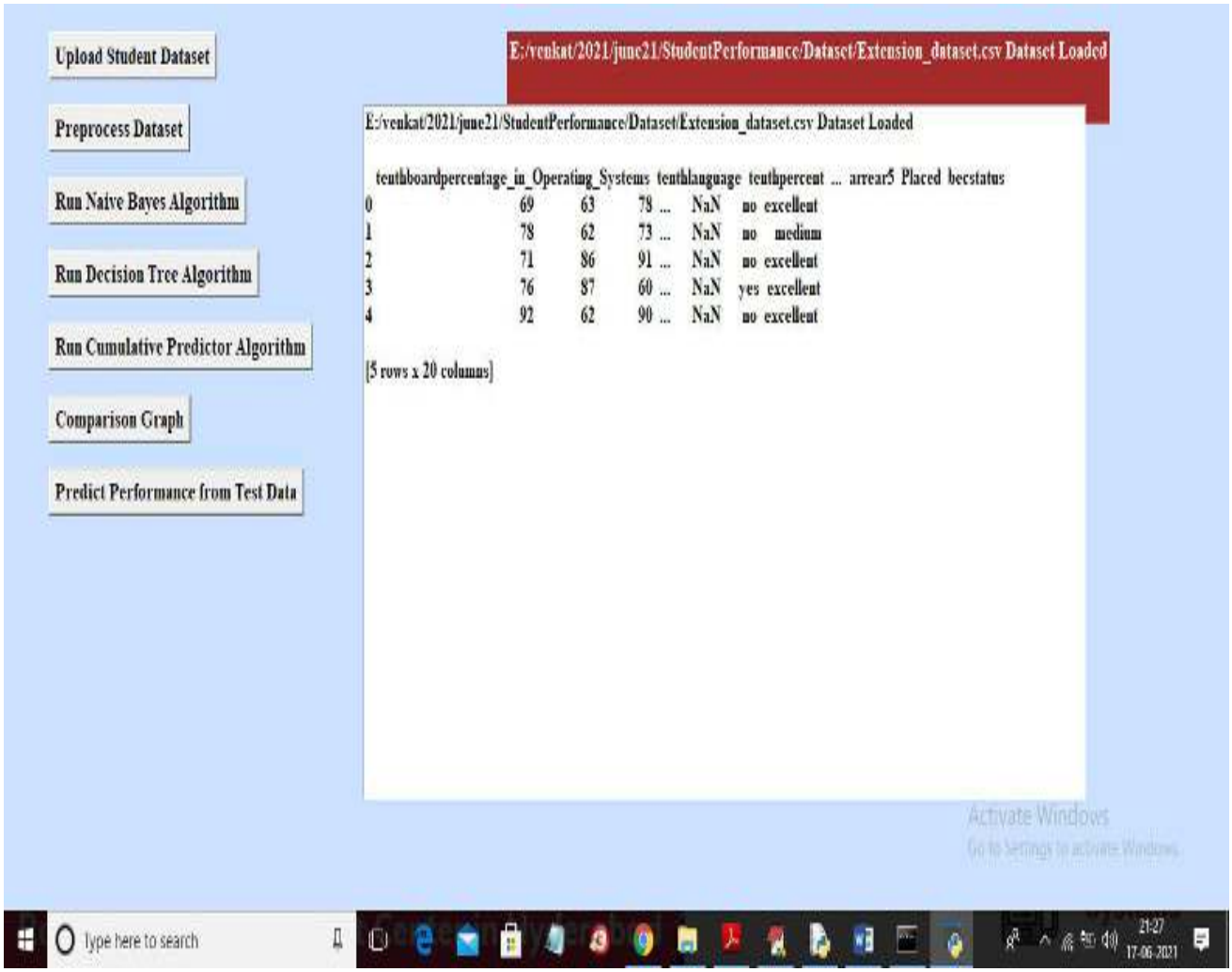


5.2 UPLOADING THE TRAINING DATA TO TRAIN THE DIFFERENT ALGORITHMS

USED.

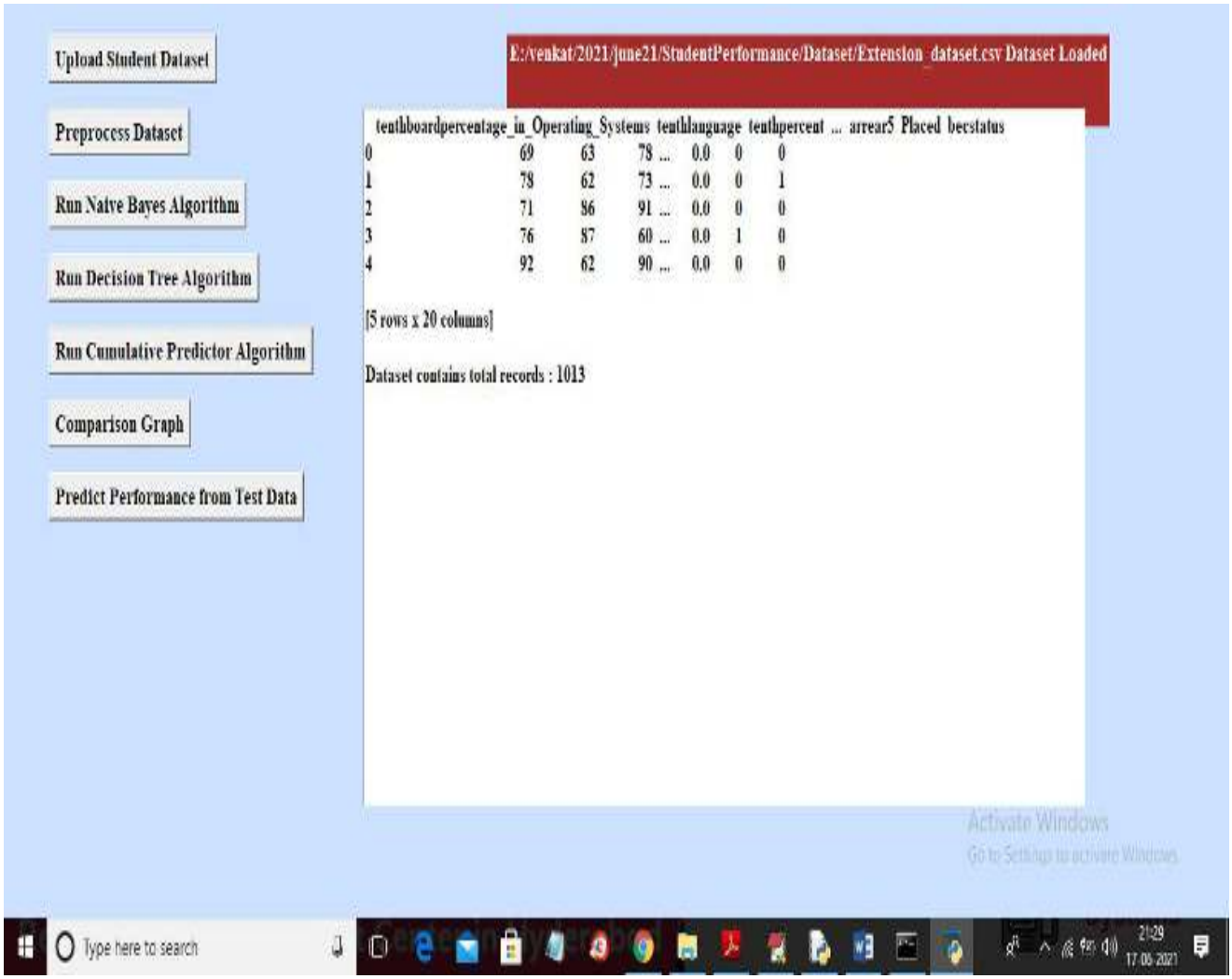


5.3 UPLOADED DATA BEFORE PREPROCESSING

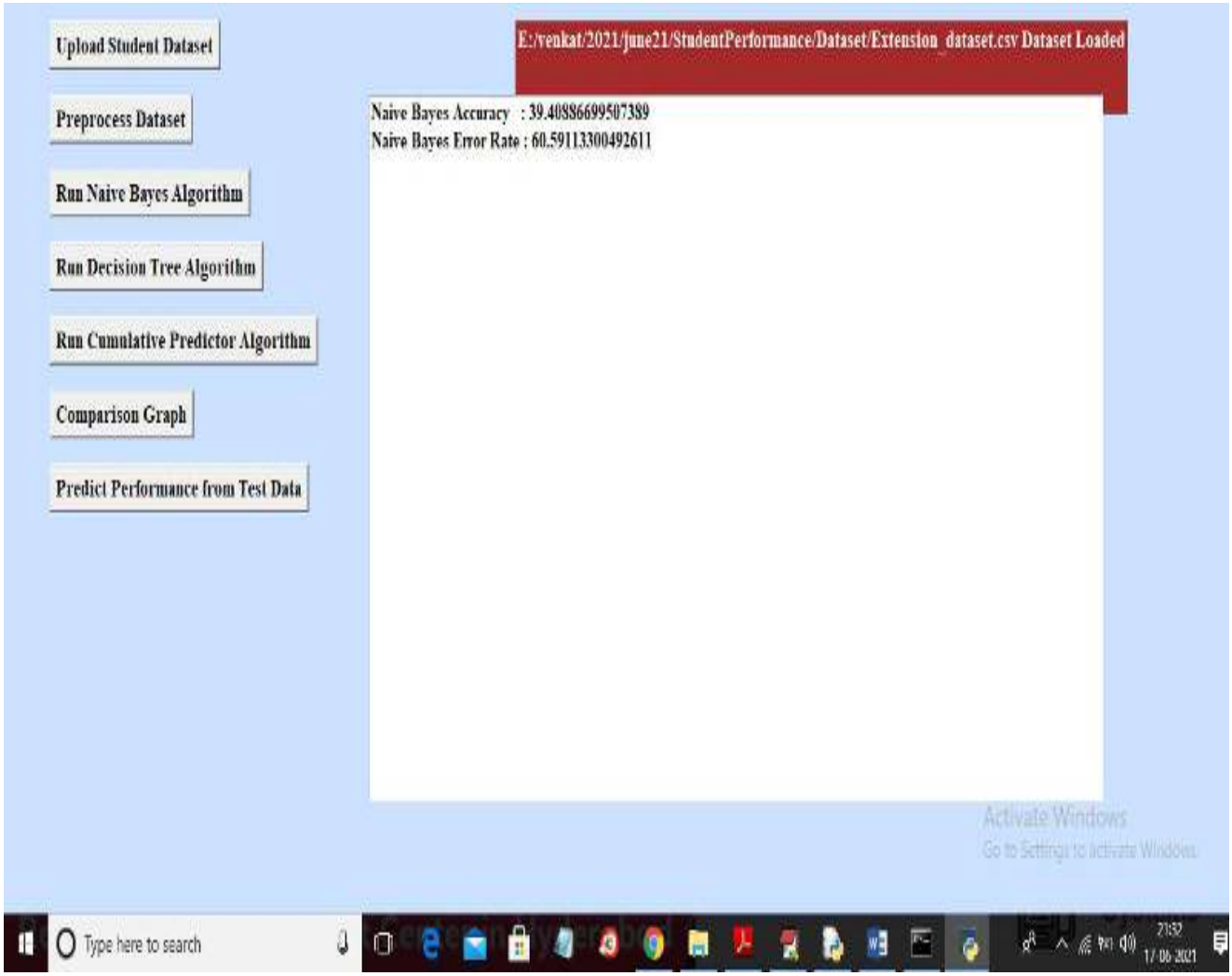


5.4 UPLOADED DATA AFTER PREPROCESSING IT

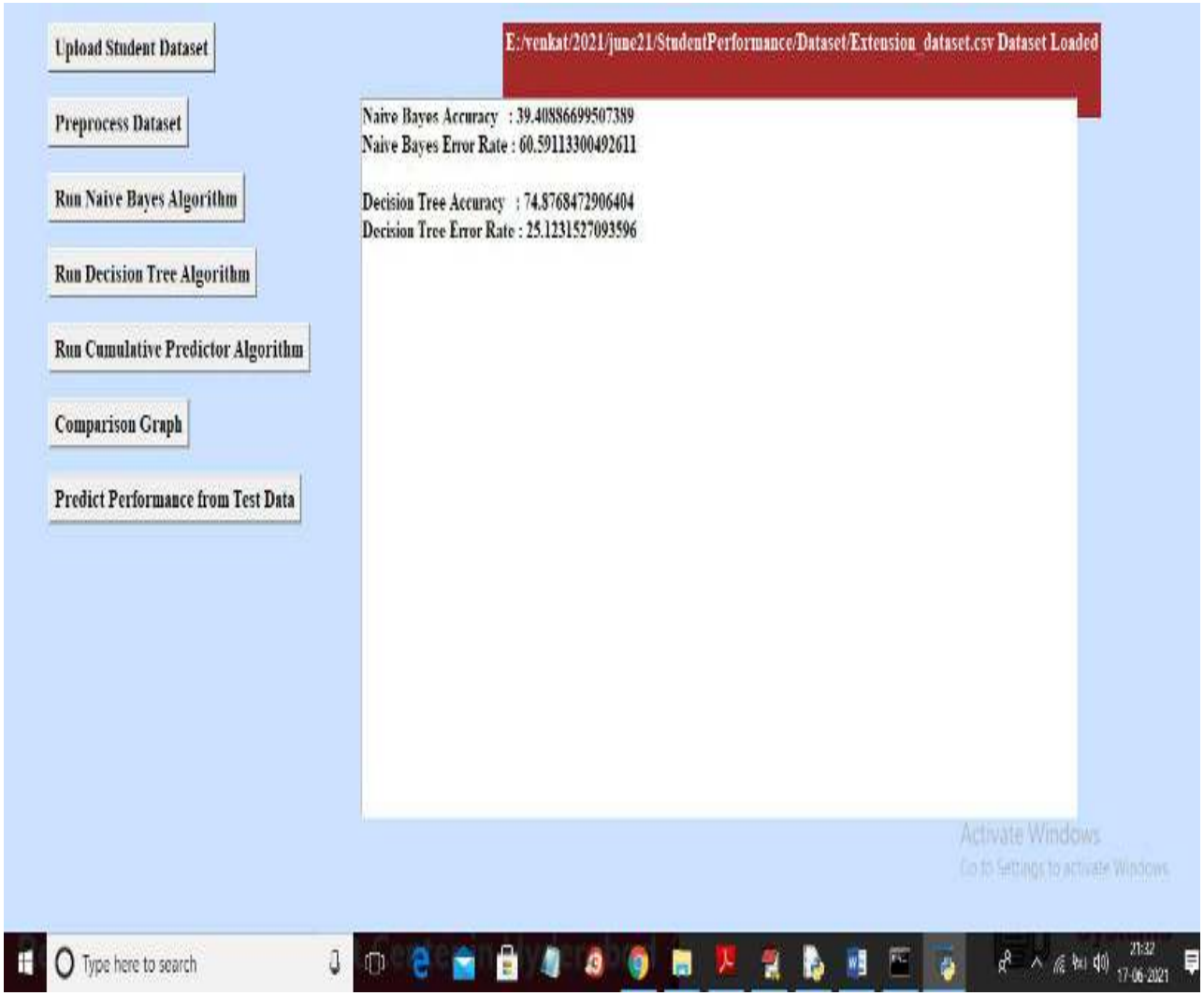
(THE NULL DATA, YES AND NO DATA ARE REPLACED BY 0 AND 1.)



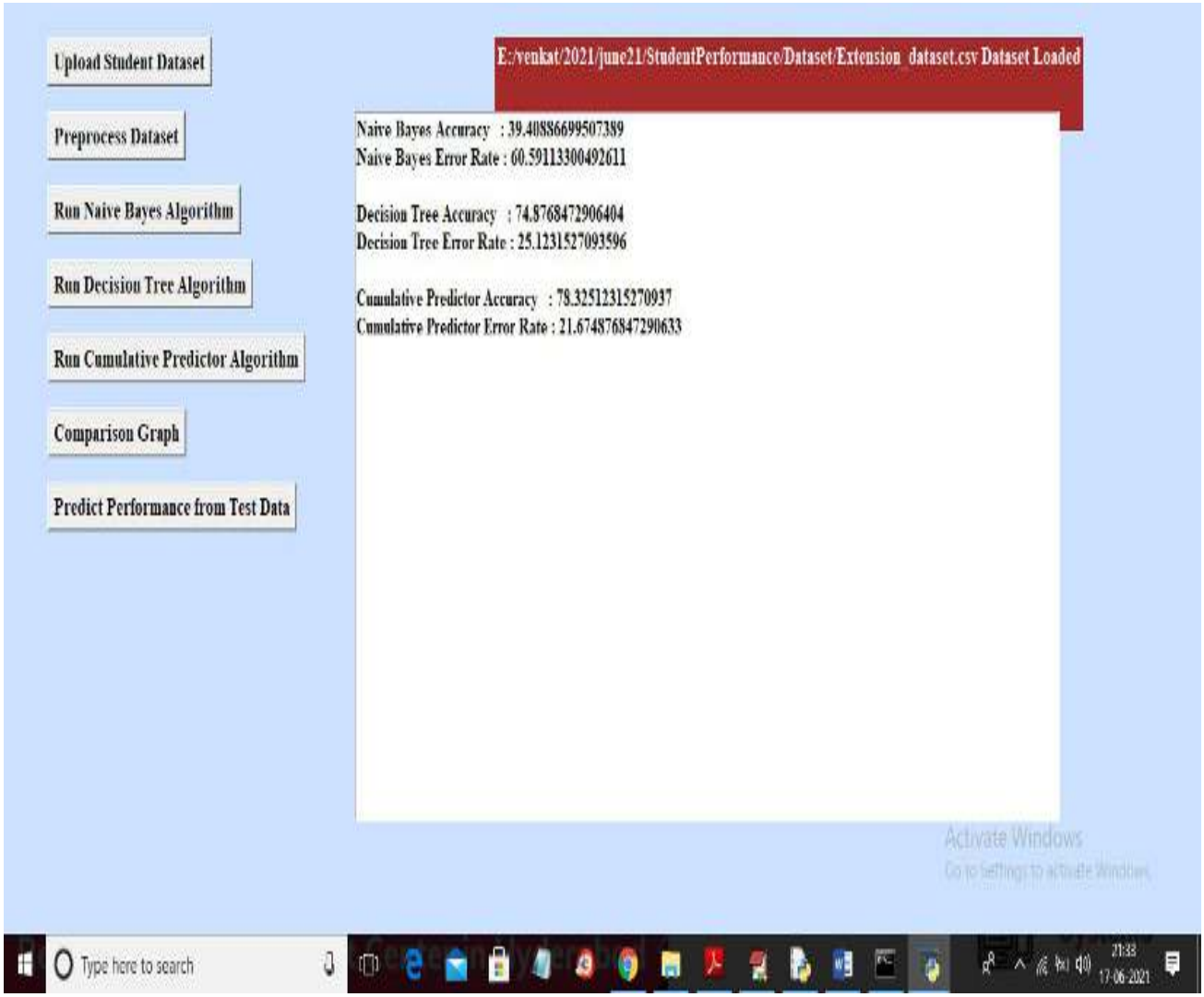
5.5 THE ACCURACY AND ERROR RATE AFTER TRAINING THE NAÏVE BAYES MODEL USING PREPROCESSED DATASET



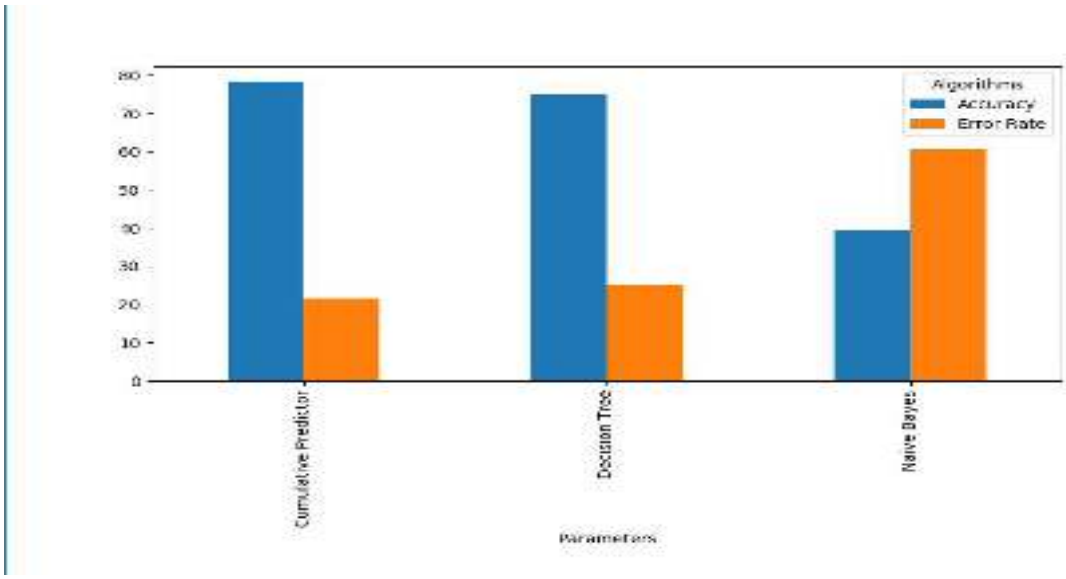
5.6 THE ACCURACY AND ERROR RATE AFTER TRAINING THE DECISION TREE MODEL USING PREPROCESSED DATASET

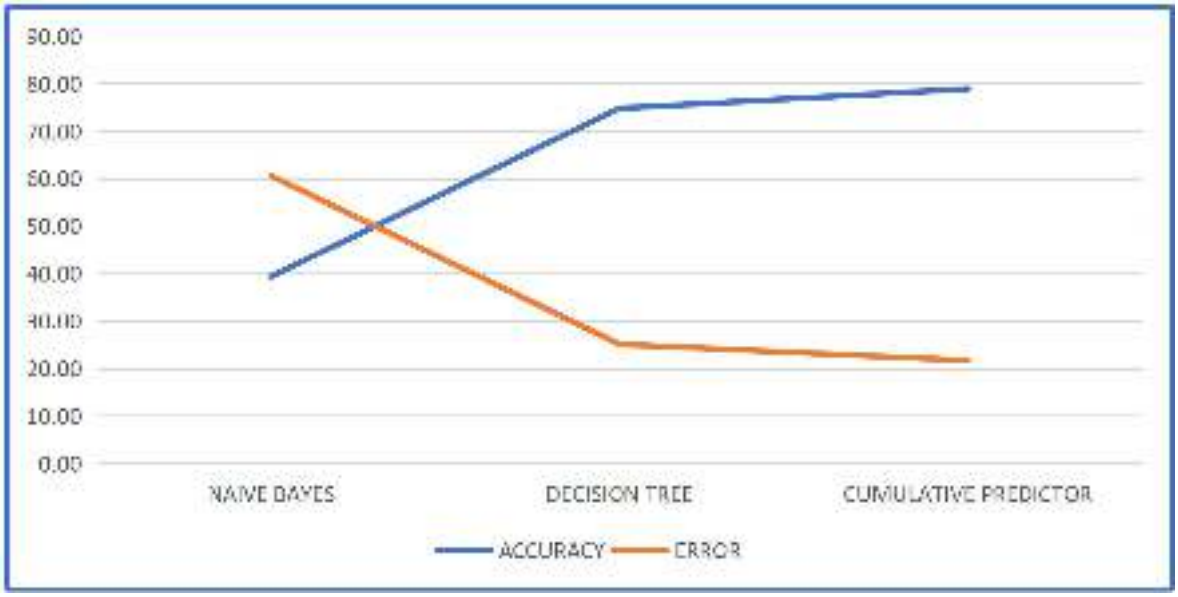


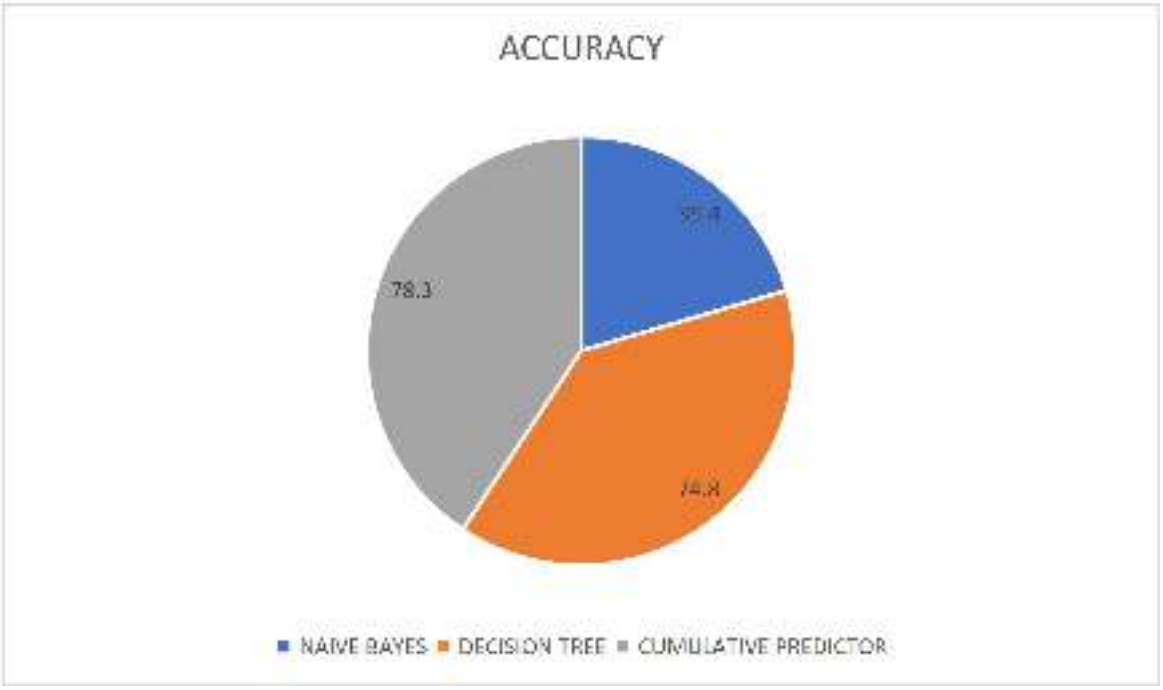
5.7 THE ACCURACY AND ERROR RATE AFTER TRAINING THE CUMULATIVE PREDICTOR MODEL USING PREPROCESSED DATASET

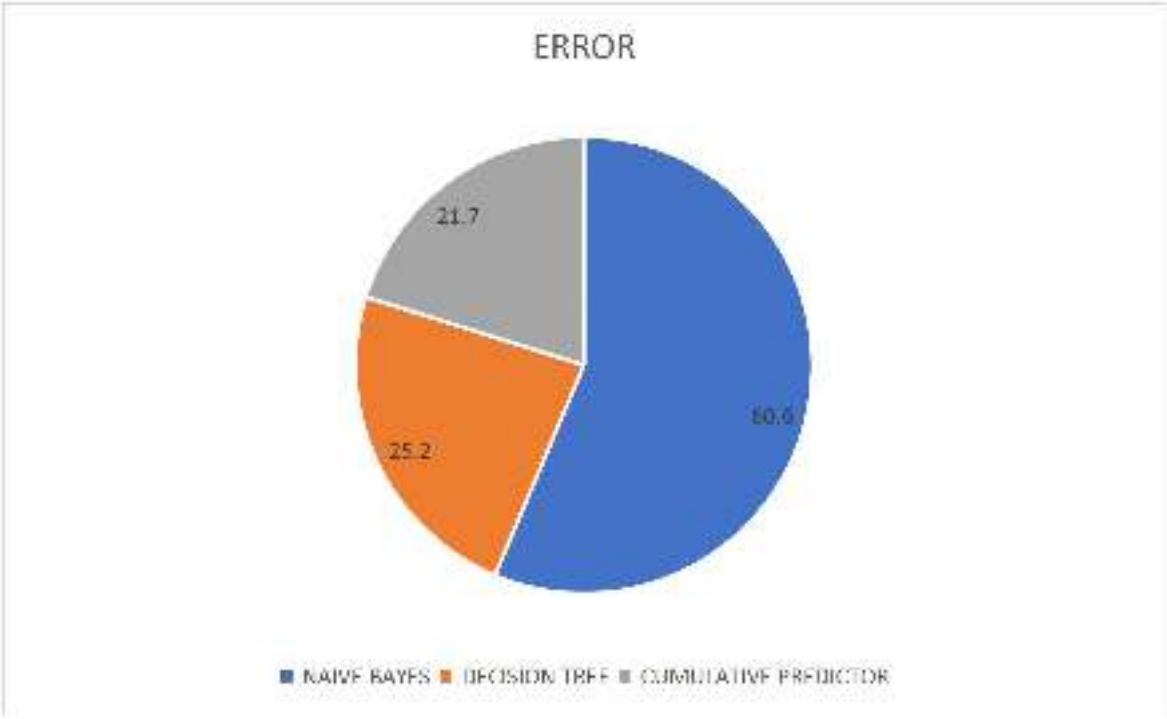


5.8 THE ACCURACY AND ERROR RATE COMPARISION GRAPH

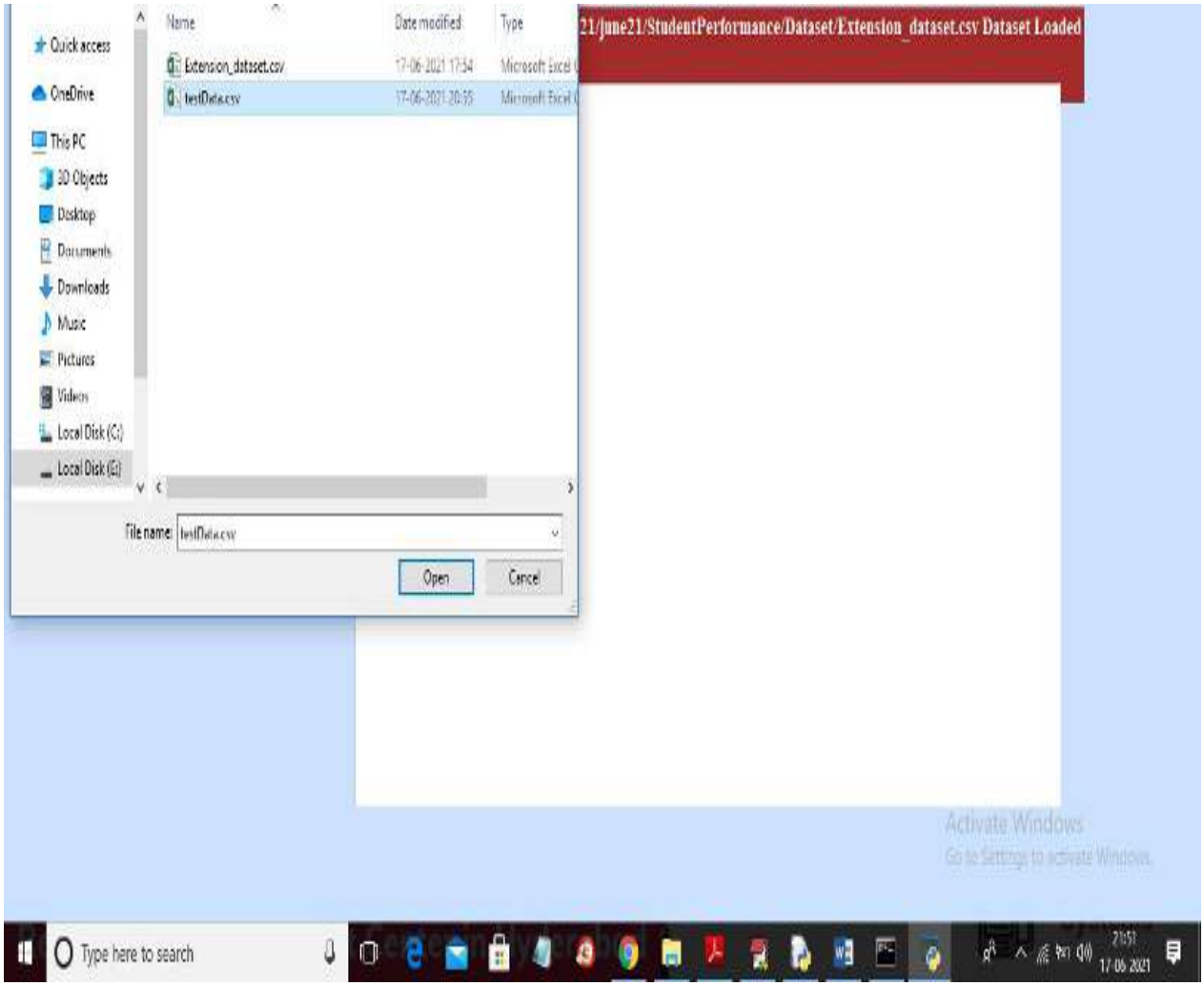




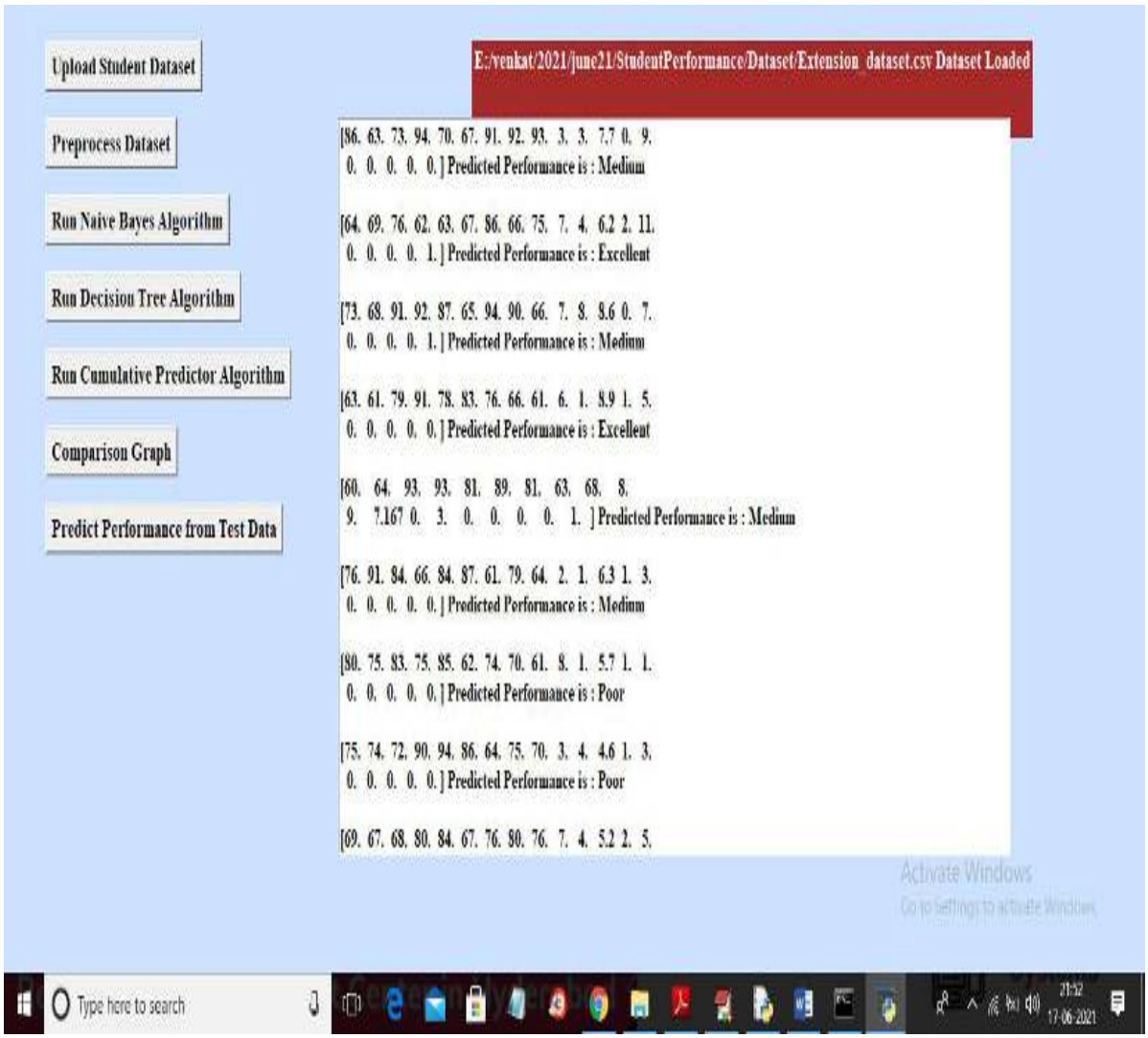

ACCURACY
78.3
NAIVE BAYES
DECISION TREE
CUMULATIVE PREDICTOR


ERROR
21.7
25.2
NAIVE BAYES
DECISION TREE
CUMULATIVE PREDICTOR

5.9 UPLOADING THE TEST DATA TO PREDICT AN OUTCOME AS EXCELLENT, MEDIUM OR POOR



5.10 PERFORMANCE PREDICTION FROM THE TEST DATA



6.TESTING

6.1 TESTING STRATEGIES

6.1.1 UNIT TESTING

Unit testing, a testing technique using which individual modules are tested to determine if there are issues by the developer himself.. it is concerned with functional correctness of the standalone modules. The main aim is to isolate each unit of the system to identify, analyze and fix the defects.

Unit Testing Techniques:

Black Box Testing - Using which the user interface, input and output are tested.

White Box Testing –Used to test each one of those functions behavior is tested.

6.1.2 DATA FLOW TESTING

Data flow testing is a family of testing strategies based on selecting paths through the program's control flow in order to explore sequence of events related to the status of Variables or data object. Dataflow Testing focuses on the points at which variables receive and the points at which these values are used.

6.1.3 INTEGRATION TESTING

Integration Testing done upon completion of unit testing, the units or modules are to be integrated which gives raise too integration testing. The purpose of integration testing is to verify the functional, performance, and reliability between the modules that are integrated.

6.1.4 BIG BANG INTEGRATION TESTING

Big Bang Integration Testing is an integration testing Strategy wherein all units are linked at once, resulting in a complete system. When this type of testing strategy is adopted, it is difficult to isolate any errors found, because attention is not paid to verifying the interfaces across individual units.

6.1.5 USER INTERFACE TESTING

User interface testing, a testing technique used to identify the presence of defects is a product/software under test by Graphical User interface [GUI].

6.2 TEST CASES

S.NO	INPUT	If available	If not available
1	Upload dataset	Dataset loaded	There is no process
2	Data preprocessing	process of transforming raw data into an understandable forma	There is no process
3	Data preprocessing	process of transforming raw data into an understandable forma	There is no process
4	Run naïve bayes	Algorithm accuracy displayed	There is no process
5	Run decision tree	Algorithm accuracy displayed	There is no process
6	Run cumulative predictor	Algorithm accuracy displayed	There is no process
7	Comparison graph	Graph displayed	There is no process

8	Predict performance from test data	we will upload test data and then cumulative predictor will predict performance	There is no process

7.CONCLUSION

7.CONCLUSION & FUTURESCOPE

This system can be very easily implemented and utilized by any educational institution. It can be used by faculties and students who do not have any knowledge on data mining techniques. Although there are so many benchmarks comparing the performance and accuracy of different classification algorithms, there are still very few experiments carried out on educational datasets. In this work, we compare the performance and the interpretation levels of the output of different classification techniques applied on educational datasets and finally develop a much more efficient algorithm called the cumulative predictor algorithm. Our experimentation shows that there is not one algorithm that obtains significantly better classification accuracy, so ensemble of classifier is created. Thus the final result of cumulative predictor algorithm is much more accurate than existing algorithms.

FUTURESCOPE

In this project, the prediction using the decision tree and Cummulative predictor generated is not updated dynamically with in the system's source codes. Thus, in future, a dynamic prediction model could be implemented by train the prediction model itself whenever a new training set are fed into the system .

8.BIBILIOGRAPHY

8.1 REFERENCES

1. Ahmad, F., Ismail, N.H. and Aziz, A.A. (2015)'The prediction of students' academic performance using classification data mining techniques', Applied Mathematical Sciences, Vol. 9, No. 129, pp.6415–6426, HIKARI Ltd.

2. García, E.P.I. and Mora, P.M. (2011) 'Model prediction of academic performance for first year students', Proceedings of 10th Mexican International Conference on Artificial Intelligence, pp.169–174.

3. Jain, R. and Minz, S. (2008) 'Drawing conclusion from forest cover type datathe hybridized rough set model', Journal of the Indian Society of Agricultural Statistics, Vol. 62, No. 1, pp.75–84.

4. Kabakchieva, D. (2013) 'Predicting student performance by using data mining methods for

5. classification', Cybernetics and Information Technologies, Vol. 13, No. 1, pp.61–72 [online]

8.2 GITHUB LINK

https://github.com/ShivaTeja13/STUDENTS-PERFORMANCE-ANALYSIS-SYSTEM