**PREDICTION AND ANALYSIS OF STUDENT PERFORMANCE IN DATA MINING**

**A PROJECT REPORT**

***Submitted by***

**V.PUSHPA (810015104069)**

**D.RESHIKA (810015104076)**

***in partial fulfillment for the award of the degree***

***of***

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**



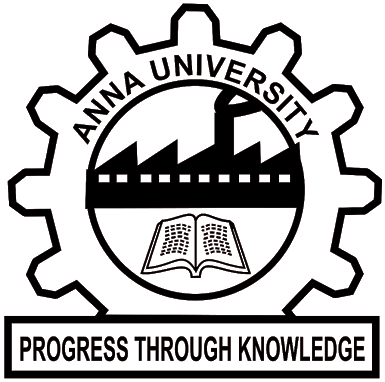
**UNIVERSITY COLLEGE OF ENGINEERING-BIT CAMPUS**

**TIRUCHIRAPPALLI-620 024**

**ANNA UNIVERSITY::CHENNAI-620 025**

**APRIL 2019**

**ANNA UNIVERSITY: CHENNAI 600 025**



**BONAFIDE CERTIFICATE**

Certified that this project report **“ PREDICTION AND ANALYSIS OF STUDENT PERFORMANCE IN DATA MINING ”** is the bonafide work of **“V.PUSHPA(810015104069)** and **D.RESHIKA(810015104076)”** who carried out project work under my supervision.

**SIGNATURE SIGNATURE**

Dr.D.VENKATESAN Dr.M.PADMA

**HEAD OF THE DEPARTMENT SUPERVISOR**

Department of CSE Teaching fellow, Dept of CSE

UCE, BIT Campus UCE,BIT Campus

Anna University, Anna University,

Tiruchirappalli -24 Tiruchirappali -24.

Submitted to University Viva-Voce Examination held on \_\_\_\_\_\_\_\_\_\_

**INTERNAL EXAMINER EXTERNAL EXAMINER**

**DECLARATION**

We hereby declare that the work entitled **“PREDICTION AND ANALYSIS OF STUDENT PERFORMANCE IN DATA MINING”** is submitted in partial fulfillment of requirement for the award of the degree in B.E., Computer Science and Engineering, University College of Engineering(BIT Campus), Tiruchirappalli, is a record of the our original work carried out by us during the academic year 2018-2019. Under the supervision and guidance of **Dr. M.PADMA**, Teaching fellow, Department of Computer Science and Engineering, University College of Engineering (BIT Campus), Tiruchirappalli. The extent and source of information are derived from the existing literature and have been indicated through the dissertation at the appropriate places. The matter embodied in this work is original and has not been submitted for the award of any other degree or diploma, either in this or any other university.

**SIGNATURE OF THE CANDIDATES**

V. PUSHPA(810015104069)

D.RESHIKA(810015104076)

I certify that the declaration made above by the candidate is true.

**SIGNATURE OF THE GUIDE**

**Dr.M.PADMA**

**ACKNOWLEDGEMENT**

We take the opportunity to extend our heartfelt thanks to our respected Dean **Dr. T. SENTHILKUMAR ,** for his support and impeccable guidance.

We are extremely grateful to the Head of the Department of CSE & IT, **Professor Dr.D.VENKATESAN**, for having encouraged and helped us throughout the course of our project. Without his supervision and feedback, it would have been really hard for us to finish our project in a timely manner. Thus, we feel deeply obliged for his support.

I wish to express my sincere gratitude to our project Co-ordinator **Mr. C.SURESHKUMAR,** Department of computer Science & Engineering,for her kindguidance and encouragement in carrying out this project work.

We are also grateful to our guide, **Dr.M.PADMA,** Teaching fellow, Department of Computer Science and Engineering,for having assisted and mentored us so diligently in the process of preparing our project. Without her persistent support and co-operation, we couldn’t have accomplished our ideas.

We thank our beloved parents and friends for their full support in the moral development of this project.

ABSTRACT

Data mining is widely used in educational field to extract information from huge amount of data. Student Information System provides interface for maintaining information right from the time of admission to till the completion of their courses. Predicting Student performance alert the learner to know about their current performance and improve them in future. The proposed system maintains not only the academic details but also details regarding extra-curricular activities, placement details and other resources. This System have faculty, student , parent and alumni details in all aspects and the various academic notifications and event organised by the college administration. To brings interaction among students, faculty parents and alumni. In this study, we develop a classification and Prediction model using C4.5 algorithm. The whole system will be available through secure, online interface embedded in college website.

**TABLE OF CONTENTS**

**CHAPTER NO TITLE PAGE NO**

**ABSTRACT**

**LIST OF FIGURES**

**LIST OF ABBREVIATIONS**

**1 INTRODUCTION**

1.1 Student information system 1

1.2 Data Mining 1

1.2.1 Knowledge Discovery Database 2

1.3 Machine Learning 3

1.4 Objective 5

**2 SYSTEM ANALYSIS**

2.1 Literature survey 6

2.2 Existing System 12

**3 SYSTEM SPECIFICATIONS**

3.1 Hardware Specifications 15

3.2 Software Specifications 15

**4 SYSTEM DESIGN**

4.1 Proposed System 16

4.1.1 System Architecture 16

4.2 Apriori Algorithm 17

4.3 Decision Tree Algorithm 18

4.4 Work Flow Diagram 19

**5 IMPLEMENTATION**

5.1 Dataset and attribute selection 20

5.2 Preprocessing 21

5.3 Filters 21

5.4 Classification 22

5.5 Prediction of result 23

5.6 Visualization 30

**6 CONCLUSION AND FUTURE WORK**

6.1 Conclusion 31

6.2 Future Enhancement 31

**APPENDICES**

A. Source code 32

B. Screenshots 43

**REFERENCES 47**

**LIST OF FIGURES**

**FIGURE NO FIGURE NAME PAGE NO**

2.2 Decision tree for programming domain

4.1.1 System Architecture

4.3 Work Flow Diagram

5.1 Dataset

5.5 Predicted result

**LIST OF ABBREVIATIONS**

**S.NO ABBREVIATION EXPLANATION**

1 UI User Interface

2 NB Naïve Bayes

3 KDD Knowledge Discovery Database

4 EDM Educational Data Mining

5 SVM Support Vector Machine

6 AOA Analysis of Algorithm

**CHAPTER 1**

**INTRODUCTION**

**1.1 STUDENT INFORMATION SYSTEM**

The main purpose of designing and implementation of a comprehensive student information system and user interface is to replace the current paper records. The system provides staff UI which can access all aspects of a student’s academic progress through a secure, online interface embedded in the colleges website. In addition to this it also provides student and parent UI, allowing users to access information. The system uses user authentication, displaying only relevant data. Each sub- system has authentication allowing authorized users to create or update data in that sub-system.

All data is stored securely on SQL servers managed by college administrator. All the authentication details like username, password etc. are sent through an email to the registered id. Previously, this records were stored on papers which had many drawbacks. This system provides a simple interface for the maintenance of student information. Achieving this is difficult using traditional method as the data is scattered, redundant and time consuming. This system focuses on proving data in an easy and intelligent manner to students, college staff and parents.

**1.2 DATA MINING**

Data Mining is defined as extracting information from huge sets of data. In other words, we can say that data mining is the procedure of mining knowledge from data. Data Mining can also be referred as Knowledge Discovery from Data (KDD). Data Mining is a multidisciplinary field, encompassing areas like information technology, machine learning, statistics, pattern recognition, data retrieval, neural networks, information based systems, artificial intelligence and data visualization.

To accomplish these tasks, data miners use one or more of the following techniques:

* **Data summarization** with statistics, including finding outliers.
* **Visualization** presenting a graphical summary of the data.
* **Clustering** of the data into natural categories.
* **Association rule** discovery defining normal activity and enabling the discovery of anomalies.
* **Classification** predicting the categories to which a particular record belongs.
  + 1. **KNOWLEDGE DISCOVERY DATABASE**

Knowledge discovery in databases (KDD) is the process of discovering useful knowledge from a collection of data. This widely used data mining technique is a process that includes data preparation and selection, data cleansing, incorporating prior knowledge on data sets and interpreting accurate solutions from the observed results.

**APPLICATION**

Our project is on Educational Data Mining (EDM) field. It has several applications. The areas of EDM are-

* + Analysis and visualization of data
  + Providing feedback for supporting instructors
  + Recommendations for students
  + Predicting student performance
  + Student modeling
  + Detecting undesirable student behaviors
  + Grouping students
  + Planning and scheduling

**1.3 MACHINE LEARNING**

Machine learning is the scientific study of algorithms an statistical model that computer systems use to effectively perform a specific task without using explicit instructions, relying on patterns and interface instead. It build a Mathematical model of sample data, known as “training data” in order to make predictions or decisions without being explicitly programmed to perform the task. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through unsupervised learning.

Machine learning tasks are classified into several broad categories:

* Supervised learning
* Unsupervised learning

**Supervised learning algorithm:**

In supervised learning, the algorithm builds a mathematical model from a set of data that contains both the inputs and output desired outputs. Classification is a basic task in data analysis that need for the construction of classifier. In supervised learning classification is there and it uses decision tree method in which its rule set is used to predict class of data. Different classification algorithms are Navie Bayes, Bayes Net, OneR and C4.5, random decision tree, random forest algorithm.

1. **C4.5:** The C4.5 algorithm used in Data mining as a Decision Tree Classifier which can be employed to generate a decision, based on sample of data.
2. **Navie Bayes Classification:** TheNavie Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes’ theorem with strong independence assumptions between the features.
3. **Random Forest:** ARandom Forest is a flexible , easy to use machine learning algorithm that produces, even without hyper parameter tuning, a great result most of time.
4. **Support Vector Machine:** SVM is binary classification algorithm.Given a set of points of 2 types in N dimensional place,SVM generates a (N-1) dimensional hyper plane to separate those points into 2 groups.

**Unsupervised learning algorithm:**

In unsupervised learning, Clustering and Association are there which are used to find hidden data in data sets to make decisions. Clustering is a technique to group similar data in one cluster to find common patterns from set of data without labels. Different clustering algorithms are Density Based, Simple K-Means, Hierarchical clustering algorithm. Association is a technique for discovering relations between variables in databases.

Common clustering algorithms include:

**1.Hierarchical clustering** builds a multilevel hierarchy of clusters by creating a cluster tree.

**2. K-Means clustering** partitions data into k distinct clusters based on distance to the centroid of a cluster.

**Association Rule Mining**

Association mining: Finding frequent patterns, associations, correlations, or casual structures among set of items or objects in transaction databases, relational databases, and other information repositories.

**Apriori Algorithm**

* + - The apriori algorithm is an influential algorithm for mining frequent item sets for Boolean association rules.
    - Aprori uses a “bottom up” approach where frequent subsets are extended one time at a time.

**Advantages of predicting student performance:**

1. It is useful to help the educators and learners improving their learning and teaching process**.**
2. Achieving this is difficult using traditional method as the data is scattered, redundant and time consuming.
3. Performance of students in their academic is a turning point for their brightest career.
4. All data is stored securely on SQL server managed by college administrator.

**1.4 OBJECTIVE**

* To provide a better understanding of what is going on in a Student’s mind.
* To improve the interaction of the student, parent, alumni with Institution.
* To explore different factor which affect performance Directly or Indirectly.

**CHAPTER 2**

**SYSTEM ANALYSIS**

**2.1 LITERATURE SURVEY**

**PAPER 1: A DATA MINING BASED SURVEY ON STUDENT PERFORMANCE EVALUATION SYSTEM**

**Authors** C.Anuradha , T.Velmurugan

A variety of Data mining techniques are available for different kind of applications. The structure of the curriculum also raises a very big impact about the development of student’s knowledge and their performance. Among the various data mining techniques, classification plays a vital role in Educational data mining. The foremost intention of this research is to traverse the data mining techniques which are used for the improvement of student’s performance and also identify the best suited structure of curriculum for the current environment. Particularly, this survey analyses about the use of classification algorithms ID3 and C4.5 for the student performance evaluation system.

**PAPER 2 : PREDICTION OF STUDENT OUTCOME USING DATA MINING TECHNIQUES**

**Authors**  R.Sumitha, E.S.Vinoth kumar

Educational Data Mining is an emerging discipline, concerned with various Approaches such as Predicting student performance, Analysis and visualization of data, Providing feedback for supporting instructors, Recommendations for students and so on that automatically extracts meaning from large repositories of data generated by or related to people's learning activities in educational setting. One of the biggest challenges is to improve the quality of the educational processes so as to enhance student’s performance. Thus, it is crucial to set new strategies and plans for a better management of the current processes. This model helps to predict student’s future learning outcomes using data sets of senior students. Thereby to design student’s data model using J48 algorithm which proved to be an efficient algorithm in terms of accuracy identified by a comparative study of data mining classification algorithms.

**PAPER 3 : PERFORMANCE PREDICTION OF STUDENTS USING DISTRIBUTED DATA MINING**

**Authors** Krina Parnar, Prof.Dineshkumar Vaghela,Dr.Priyanka Sharma

The performance of students in higher education in India is a turning point in the academics for all students for their brightest career. In today’s generation the amount of data stored in educational database increasing at a great rate. These databases contain secret information for improvement of students’ performance; these data can be located at different nodes in distributed system. Classification and prediction are among the major techniques in Data mining and widely used in various fields. Data mining methods are often implemented at many advance universities today for analyzing available data and extracting information and knowledge to support decision- making. 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.

**PAPER 4: PREDICTING STUDENT PERFORMANCE IN EDM**

**Author** Bo Guo, Rui Zhang,Guang Zu,Chuangming Shi,Li Yang

Predicting student academic performance has been an important research topic in Educational Data Mining (EDM) which uses machine learning and data mining techniques to explore data from educational settings. However measuring academic performance of students is challenging since students academic performance hinges on diverse factors. The interre- lationship between variables and factors for predicting performance participate in complicated nonlinear ways. Traditional data mining and machine learning techniques may not be applied directly to these types of data and problems. In this study we develop a classiﬁcation model to predict student performance using Deep Learning which automatically learns multiple levels of representation. We pre-train hidden layers of features layerwisely using an unsupervised learning algorithm sparse auto-encoder from unlabeled data, and then use supervised training for ﬁne- tuning the parameters. We train model on a relatively large real world students dataset, and the experimental results show the effectiveness of the proposed method which can be applied into academic pre-warning mechanism.

**PAPER 5: STUDENT PERFORMANCE PREDICTION BY USING DATA MINING CLASSIFICATION ALGORITHM**

**Authors** Dorina Kabakchieva

This paper presents the results from data mining research, performed at one of the famous and prestigious Bulgarian universities, with the main goal to reveal the high potential of data mining applications for university management and to contribute to more efficient university enrolment campaigns and to attracting the most desirable students. The research is focused on the development of data mining models for predicting student performance, based on their personal, pre- university and university-performance characteristics. The dataset used for the research purposes includes data about students admitted to the university in three consecutive years. Several well known data mining classification algorithms, including a rule learner, a decision tree classifier, a neural network and a Nearest Neighbour classifier, are applied on the dataset. The performance of these algorithms is analyzed and compared.

**PAPER 6 : ON IMPROVING STUDENT PERFORMANCE PREDICTION IN EDUCATION SYSTEM USING ENHANCED DATA MINING TECHNIQUE**

**Authors** Dr. K. Karthikeyan ,P.Kavipriya

This paper focused on improving student performance prediction, based on their personal and academic performance characteristics. Due to the incredible growth in recent technology like social media, it may deter the students from their actual track, and this is one of the reasons for the students to perform poor in academic activities and it even leads to course drop outs. Predicting students’ performance will alert the learner to know about their performance and it gives as a chance to improve their performance in future. The dataset used for the research purposes includes data about students’ performance from the academic and other class room activities in the college during the course time. , Educational data mining algorithms is used to predict the student performance which is a module in automated intelligent education systems

**PAPER 7 : ANALYSIS OF STUDENT PERFORMANCE BY USING DIFFERENT DATA MINING CLASSIFIERS**

**Author**  Hilal Almarabeh

Data mining is the analysis of a large dataset to discover patterns and use those patterns to predict the likelihood of the future events. Data mining is becoming a very important field in educational sectors and it holds great potential for the schools and universities. There are many data mining classification techniques with different levels of accuracy. The objective of this paper is to analyze and evaluate the university students' performance by applying different data mining classification techniques by using WEKA tool. The highest accuracy of classifier algorithms depends on the size and nature of the data. Five classifiers are used NaiveBayes, Bayesian Network, ID3, J48 and Neural Network Different performance measures are used to compare the results between these classifiers. The results shows that Bayesian Network classifier has the highest accuracy among the other classifiers.

**PAPER 8: PREDICTION OF STUDENT PERFORMANCE USING WEKA TOOL**

**Author** Gurmeet Kaur, Williamjit Singh

Data mining is widely used in educational field to find the problems arise in this field. Student performance is of great concern in the educational institutes where several factors may affect the performance. For prediction the three required components are: Parameters which affect the student performance, Data mining methods and third one is data mining tool. These Parameters may be psychological, personal, environmental. We conduct this study to maintain the education quality of institute by minimizing the diverse affect of these factors on student’s performance. In this Paper, Prediction of student Performance is done by applying Naïve bayes and J48 decision tree classification techniques WEKA tool. By applying data mining techniques on student data we can obtain knowledge which describes the student performance. This knowledge will help to improve the education quality, student’s performance and to decrease failure rate. All these will help to improve the quality of institute.

**PAPER 9:CLASSIFICATION AND PREDICTION BASED DATA MINING ALGORITHM TO PREDICT SLOW LEARNER IN EDUCATION SECTOR**

**Author** Parneet Kaura,Manpreet Singhb,Gurpreet Singh Josanc

Educational Data Mining field concentrate on Prediction more often as compare to generate exact results for future purpose. In order to keep a check on the changes occurring in curriculum patterns, a regular analysis is must of educational databases. This paper focus on identifying the slow learners among students and displaying it by a predictive data mining model using classification based algorithms. Real World data set from a high school is taken and filtration of desired potential variables is done using WEKA an Open Source Tool. The dataset of student academic records is tested and applied on various classification algorithms such as Multilayer Perception, Naïve Bayes, SMO, J48 and REPTree using WEKA an Open source tool. As a result, statistics are generated based on all classification algorithms and comparison of all five classifiers is also done in order to predict the accuracy and to find the best performing classification algorithm among all. In this paper, a knowledge flow model is also shown among all five classifiers. This paper showcases the importance of Prediction and Classification based data mining algorithms in the field of education and also presents some promising future lines.

**PAPER 10: A HIGHER EDUCATION PREDICTIVE MODEL USING DATA MINING TECHNIQUES**

**Author** Subha laxmi Panda, Tripti Swarnkar

The main objective of the higher educational organization is to provide high quality and necessary education to its students. The two goals of data mining in Indian education system is to analyze and enhance the chronicle way of recent educational data mining advances development; the second is to preserve, organize and discuss the content of the result which is produced by a data mining approach. The use of various data mining techniques such as random forest, decision tree, etc in Indian education processes will help to improve students' performance and provide a broad decision management skill in selection of courses as per their retention rate. This paper focuses on the model representation for analyzing the different data mining techniques in an Indian education system. Also the paper reviews a comparative study of ID3, K-Means, Naïve Bayes, Random Forest algorithm. In this paper, we have proposed the approach of Random Forest to predict the career decision for the 12th passing out students. The use of Random Forest has helped the students to take a correct appropriate decision as per their interest and skills and acts a career counselor toolbox.

**2.2 EXISTING SYSTEM**

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

* Poor
* Average
* excellent

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

**2.2.1 Information gathering**

The data set for study has been collected from examination cell. The study considers the academic performance of students from semester 1 to semester 7 of B.E program.

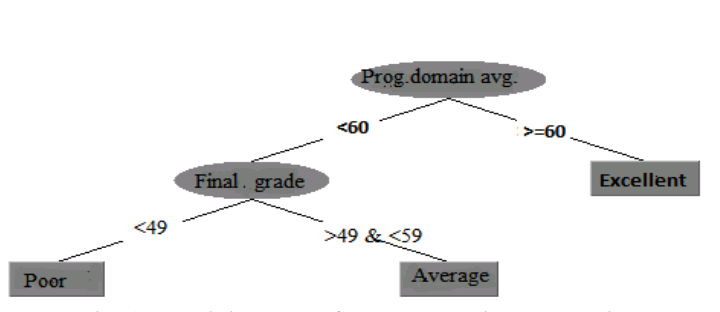
**2.2.2 Preprocessing**

1. Selection of Attributes: There may be many attributes available but all the attribute are not useful for prediction. So only the necessary attributes are taken into consideration to decide the values.

2. Deciding values for Attributes: It is necessary to decide values for attributes to avoid continuous data. This can be done by using discrete data. E.g. if average marks of the domain is from 32 to 49 then Poor. If between 50 to 59 then average. If between 60 to 80 then excellent.

**2.2.3 Interpretation**

The prediction of the result can be done for all the domain.Prediction for Programming Domain: The result shown here is for programming domain. Taking into consideration the marks obtained in programming domain subjects like Structured Programming approach (spa), OOPM, data structures, analysis of algorithm (AOA) from different semesters the prediction of the student performance in that particular domain can be done. The decision tree for programming domain is shown in the Fig.2.2.



**Fig.2.2 Decision tree for programming domain**

**CHAPTER 3**

**SYSTEM SPECIFICATIONS**

**3.1 HARDWARE SPECIFICATIONS**

Hardware is the physical components of the computer like microprocessor, hard disks, RAM and motherboard. Hardware devices are the executors of the commands provided by software applications. Computer hardware as the electronic, magnetic, and electronic devices that carry out the computing functions.

* Platform: Windows 8
* Processor: INTEL Core i3 and above
* RAM Capacity: 128 MB RAM
* Hard disk: 40 Gb

**3.2 SOFTWARE SPECIFICATIONS**

Software includes all the various forms and roles that digitally stored data may have and play in a computer (or similar system), regardless of whether the data is used as code for a CPU, or other interpreter. Software thus encompasses a wide array of products that may be developed using different techniques such as ordinary programming languages, scripting languages and etc.

* Operating System : Windows 8
* Tool : Weka 3.9

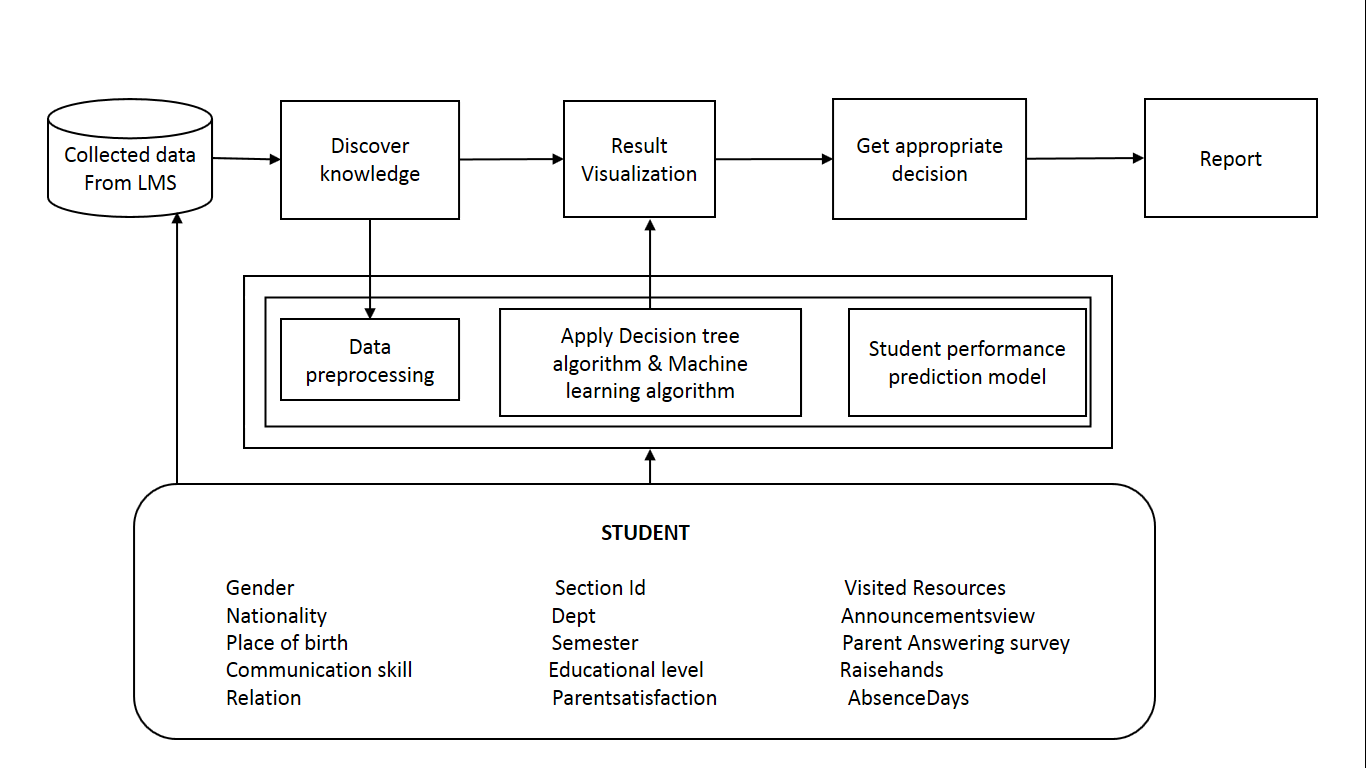
**CHAPTER 4**

**SYSTEM DESIGN**

**4.1 PROPOSED SYSTEM**

In our system, WEKA data mining software is used for the prediction of final student mark based on parameters in the given dataset. The dataset contains information about students extra-curricular skills and provide a suggestion on communication and technical skills development by which students can be built in proffessional aspects of talents.It also brings Connectivity among alumni and College.

**4.1.1 System Architecture**

****

**Fig.4.4.1.System Architecture**

Data mining provides many tasks that could be used to study the student performance. Our work will be divided into two main parts- one is prediction by classification and another one is association rule mining by using the machine learning tool ‘WEKA’. At first we will select our dataset and then perform preprocessing of it. After preprocess we will do classification over the dataset and perform prediction of result. Then we will apply association rule mining technique over the dataset and generate some rules which will be analyzed later. At last both result of prediction and association will be visualized by ‘Knowledge Flow Representation’.

In this Paper, Prediction of student Performance is done by applying Apriori classification techniques WEKA tool. By applying data mining techniques on student data we can obtain knowledge which describes the student performance. This knowledge will help to improve the education quality, student’s performance and to decrease failure rate. All these will help to improve the quality of institute.

**4.2 Apriori algorithm**

Apriori is an algorithm for frequent item set mining and association rule learning over transactional database .It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database.

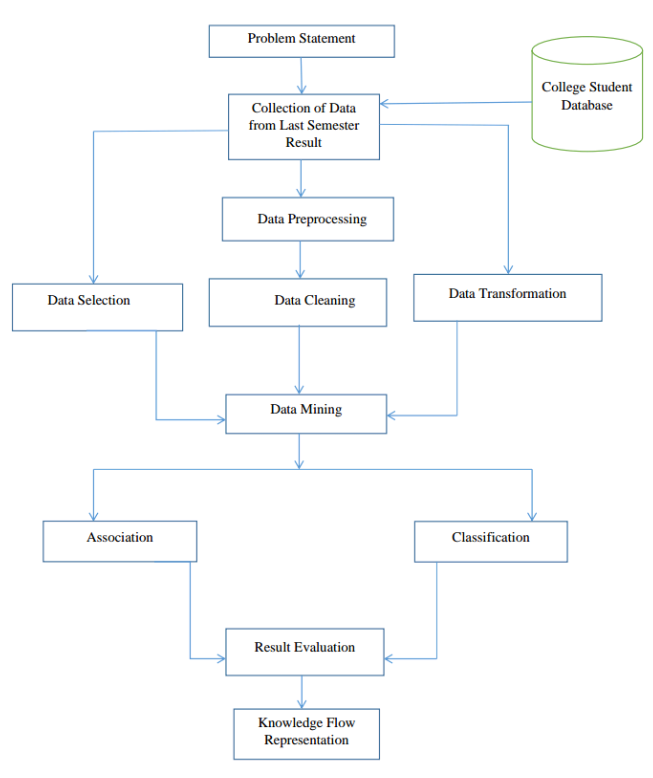
**Algorithm:**

* All itemsets having support factor greater than or equal to ,the user specified minimum support.
* All rules are having the confidence factor more significant than or similar to the user specified minimum confidence.

**4.3 DECISION TREE ALGORITHM**

1. Create a node N.
2. If all the tuples in the partition are of the same class then return N as a leaf node labeled with that class.
3. If attributes list is empty then return N as a leaf node labeled with the most common class in samples.
4. Identify the splitting attribute so that resulting partitions at each branch are as pure as possible.
5. Label node N with splitting criterion which serves as test at that node.
6. If splitting attribute is discrete valued then remove splitting attribute from attribute list
7. Let Pi be the partitions created based on the i outcomes on splitting criterion.
8. If any Pi is empty then attach a leaf with the majority class in the partition to node N.
9. Else recursively apply the complete process on each partition.
10. Return N

**4.3 WORK FLOW DIAGRAM**

****

**Fig.4.3.Work Flow Diagram**

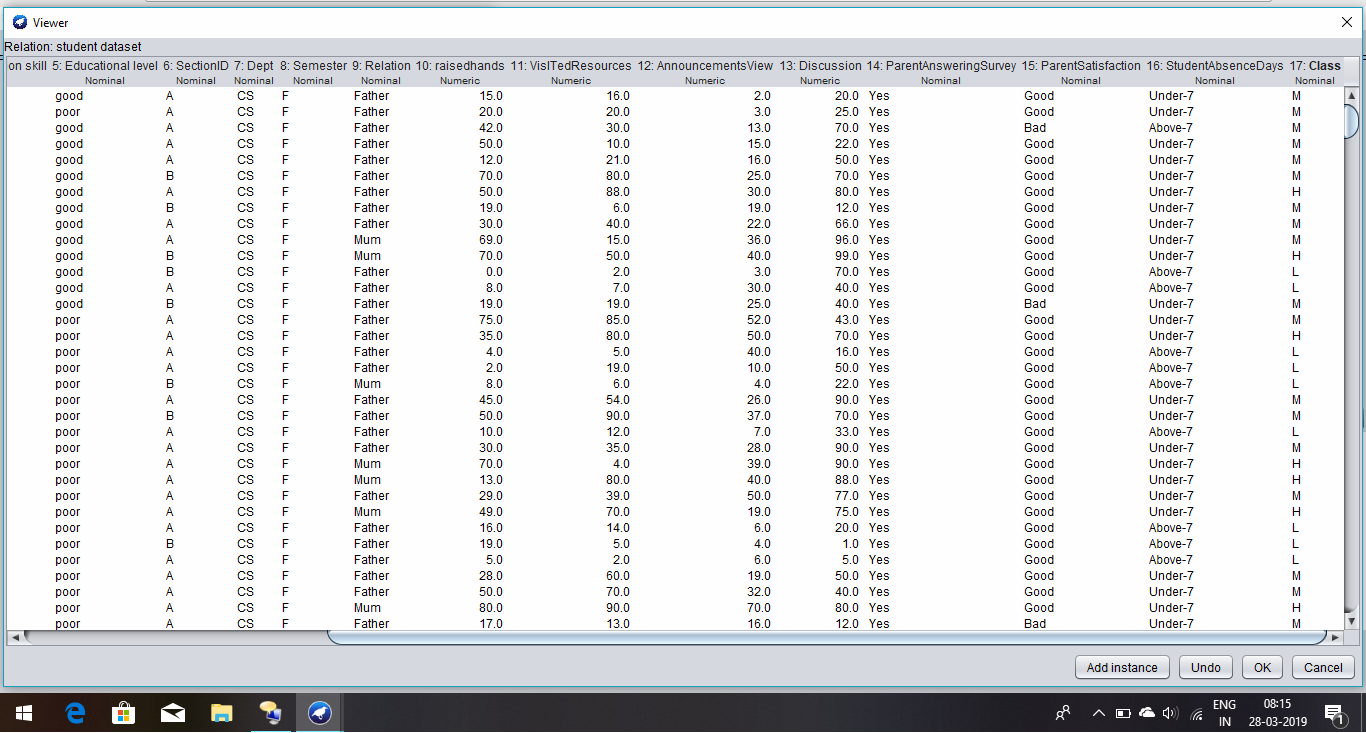
**CHAPTER 5**

**IMPLEMENTATION**

**5.1 Dataset and attribute selection**

We have collected a dummy dataset contains the result of students of last semester. The dataset contains 481 instances and 15 attributes. It has some missing values also. The data file has to be in either in ‘CSV’ format or ‘ARFF’ format.

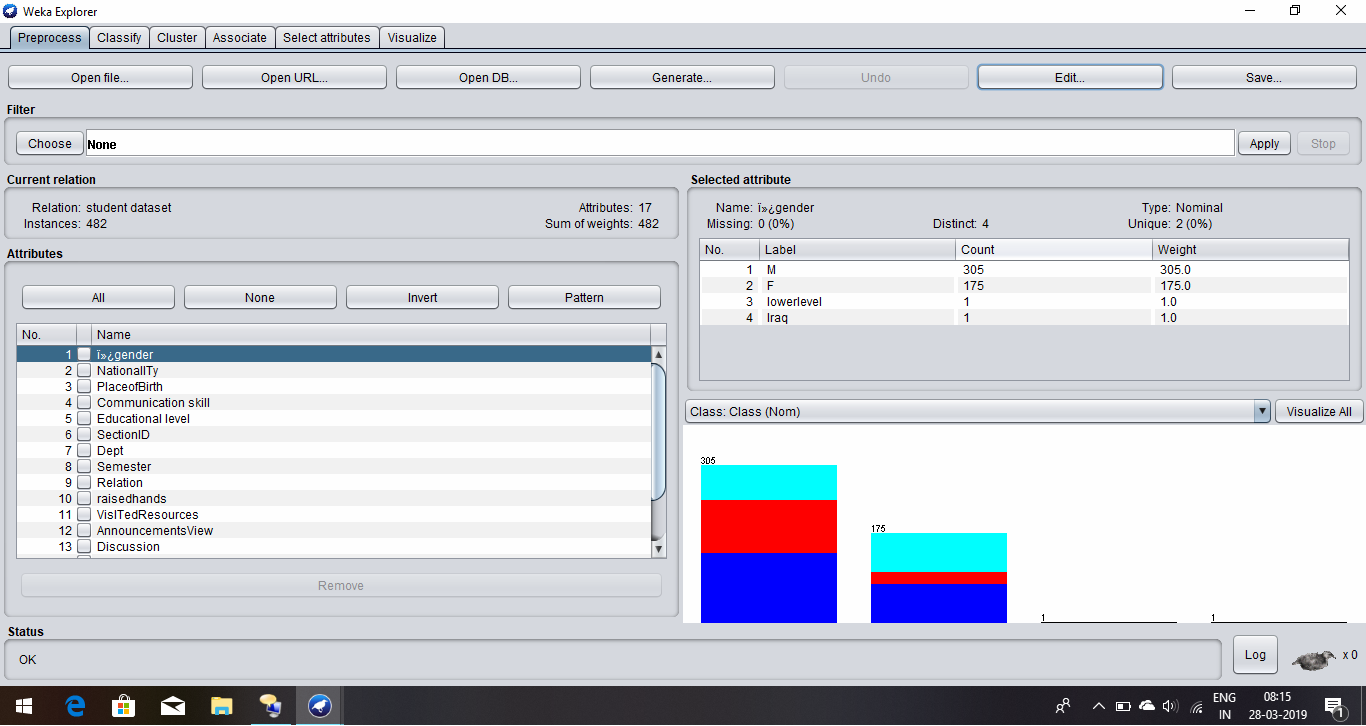
Here is the sample of our dataset which is in ‘CSV’ format.



**Fig.5.1.Dataset**

**5.2** **Preprocessing**

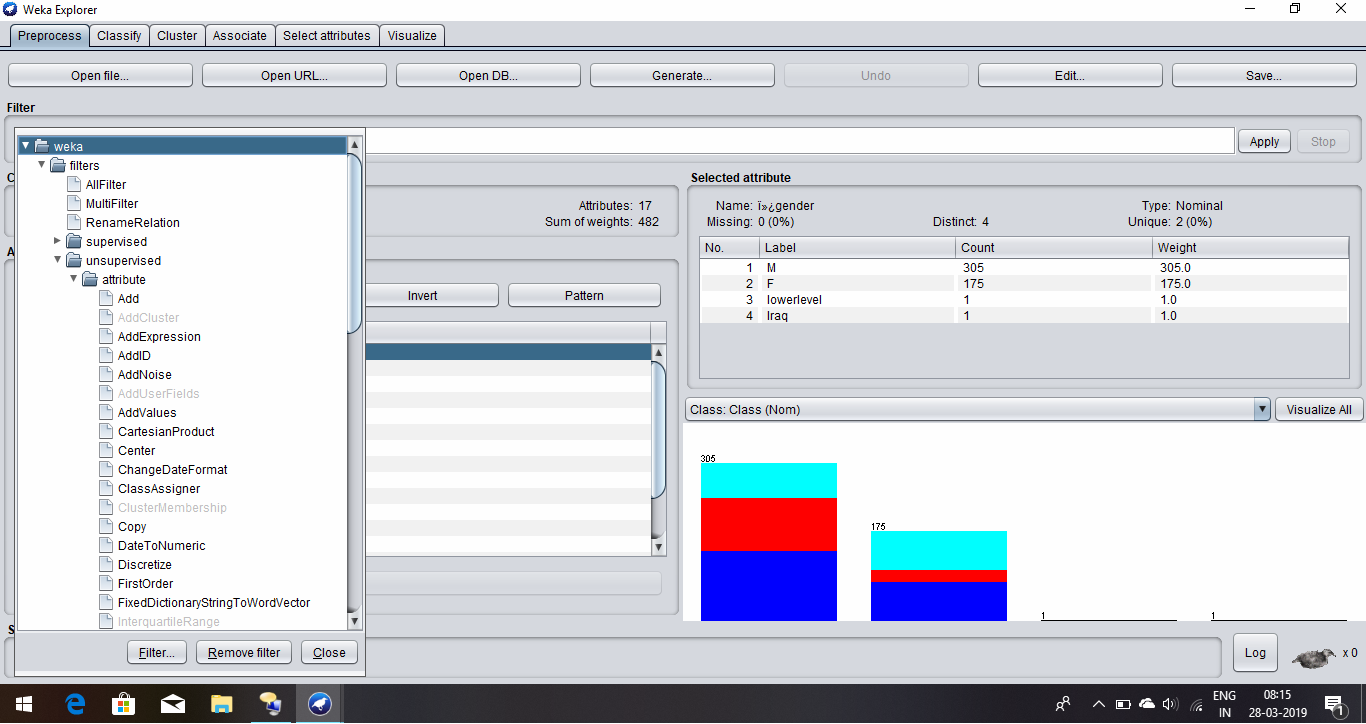
Data Preprocessing is the first step of evaluation of this project. For our project we will choose WEKA Explorer interface. Here the source data file is selected from local machine. After loading the data in Explorer, we can refine the data by selecting different options which is known as ‘Data Cleaning’ and can also select or remove attributes as per our need. The following is the preprocessed of our dataset. Left hand side of the above screen shows detail of relation name, number of attributes and number of records. Right hand side gives details of attribute values, type, and number of distinct values. Specification of every attribute is displayed in the right bottom of the screen.



**Fig.5.2.Preprocessing**

**5.3 Filters**

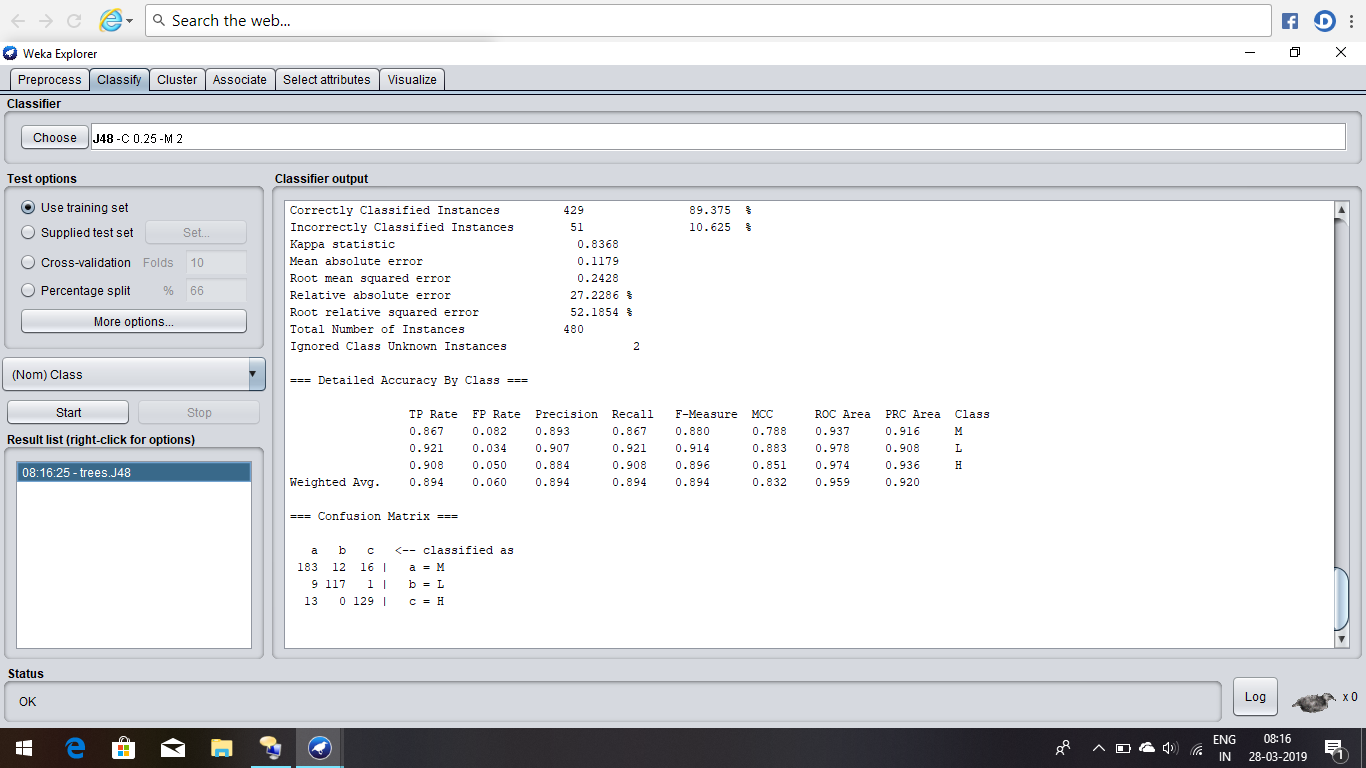
The preprocess section allows filters to be defined that transform the data in various ways. The Filter box is used to set up the filters that are required. There are mainly two categories of filters-Supervised and Unsupervised. Here we will choose unsupervised category filters. In case if the dataset is contained with any numeric values we have to covert it nominal values( as Association in WEKA can only support nominal values) by using ‘Numeric To Nominal’ filter under attribute section of Unsupervised filters. Another one filter we will apply named as ‘Replace Missing Values’ which will replace all missing values of our dataset and will make the dataset able to perform ‘Approximate Association Rule Generation’ about which we will talk later on this paper.



**Fig.5.3.Filters**

**5.4 Classification**

To predict nominal or numeric quantities we have classifiers in WEKA. For our prediction purpose we have to choose a classifier. We will select a standard classifier named as J48 for classification.



**Fig.5.4.Classification using training set**

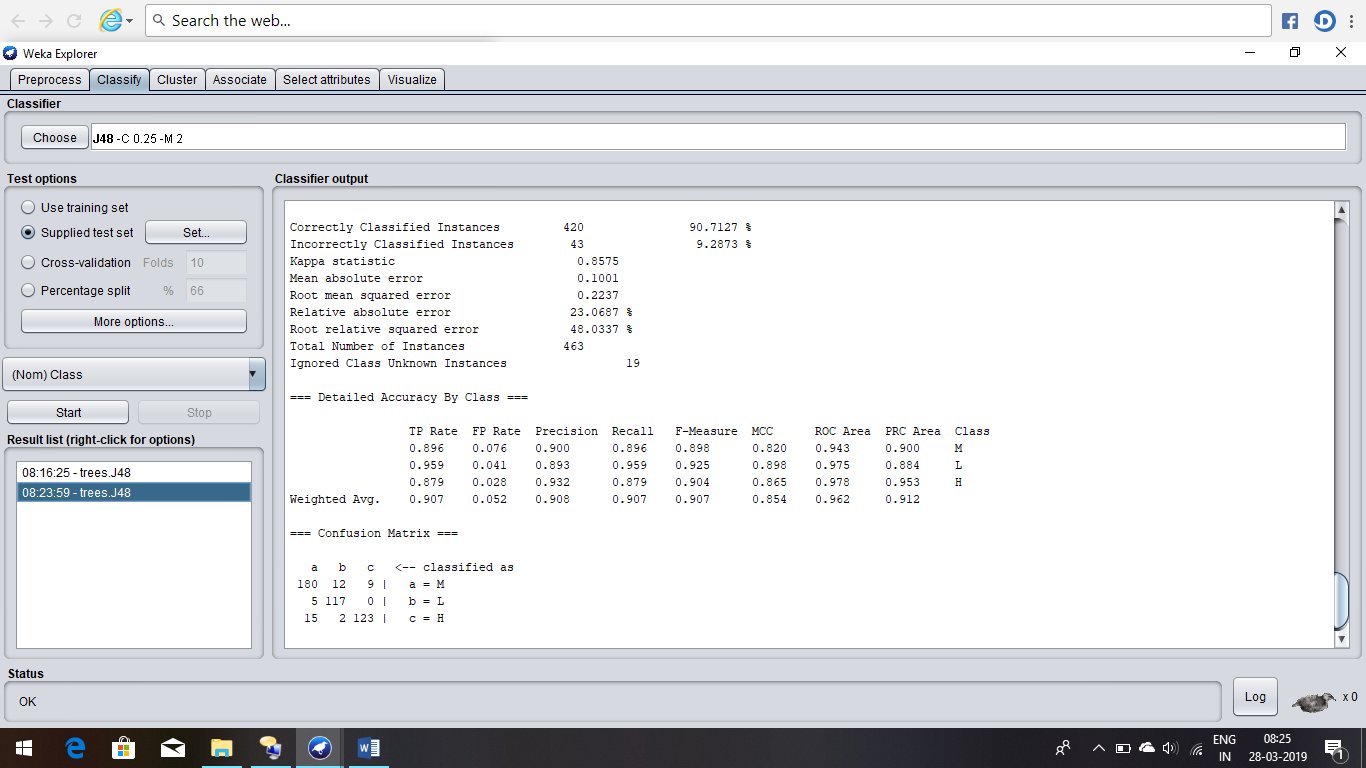
From the above example we can say J48 is a good classifier as it gives an accuracy of 89.375 % because the percentage of correctly classified instances is often called accuracy or sample accuracy. The correctly and incorrectly classified instances show the percentage of test instances that were correctly and incorrectly classified. The raw numbers are shown in the confusion matrix, with a,b,c and d representing the class labels.

Here are some others factor in classifier output,

**TP Rate**: rate of true positives (instances correctly classified as a given class)  **FP Rate**: rate of false positives (instances falsely classified as a given class) ** Precision**: proportion of instances that are truly of a class divided by the total instances classified as that class.

**Recall:** proportion of instances classified as a given class divided by the actual total in that class (equivalent to TP rate).

**F-Measure:** A combined measure for precision and recall calculated as 2 \* Precision \* Recall / (Precision + Recall)

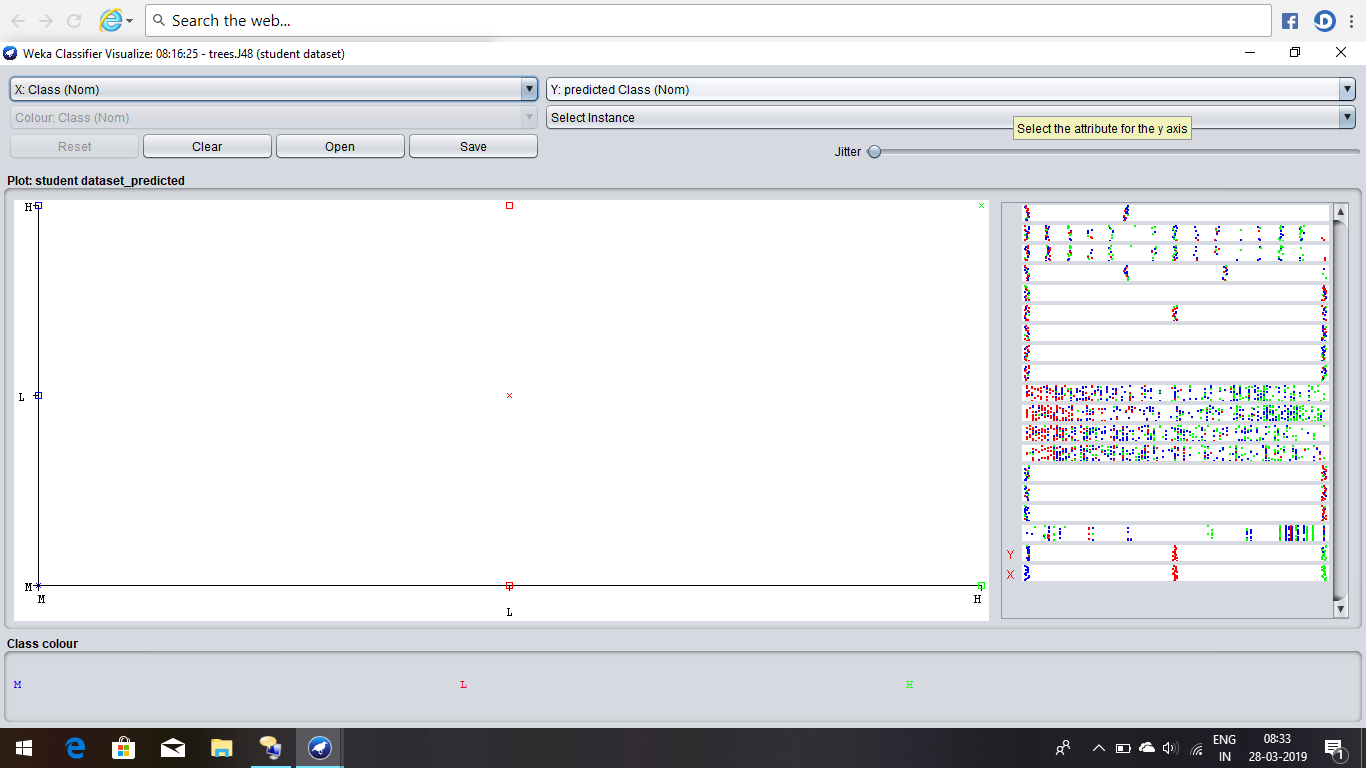


**Fig.5.4.1.Classification using test set**

Now after loading the test data we have to perform the classification by C4.5 classifier on the test dataset.

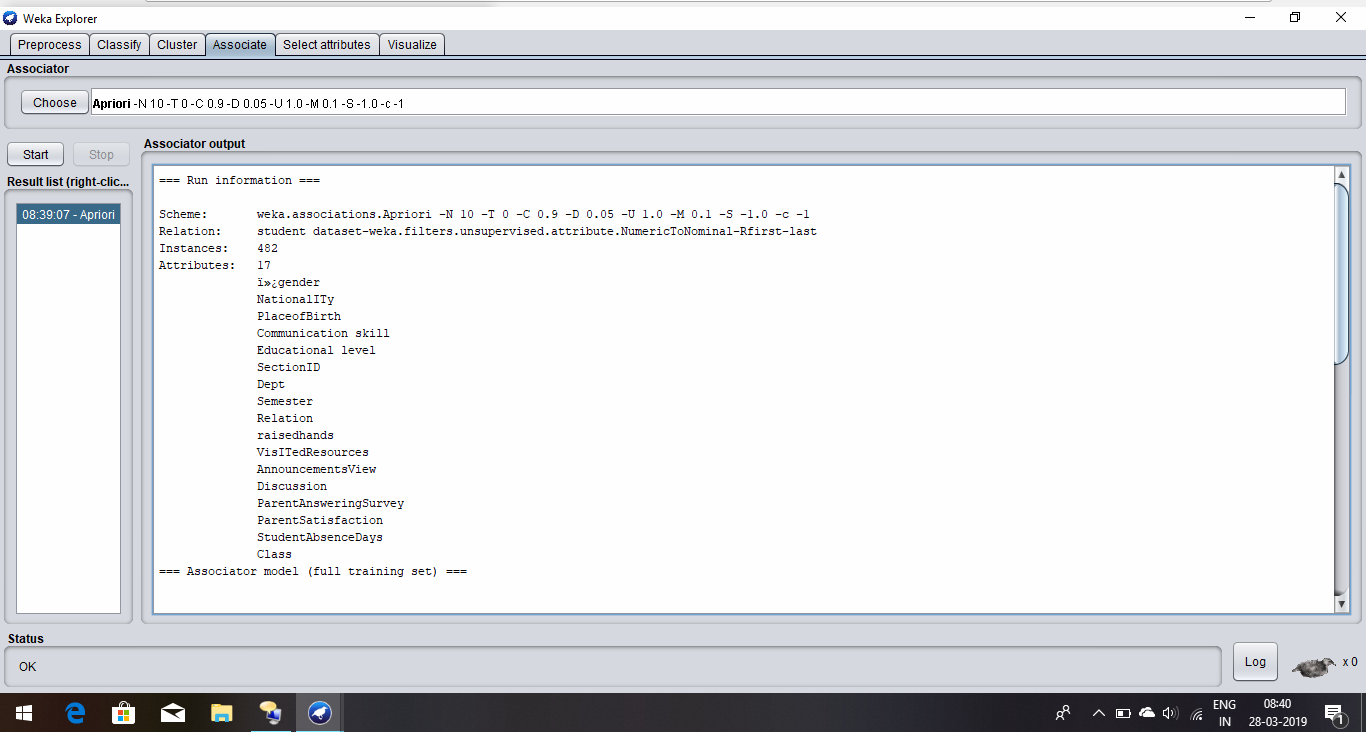
Now after classification of Test data in result list we have to select ‘Visualize Classifier errors .It will show a graph known as visualization of WEKA classifier. Then we have to save the visualization result which will be in ARFF format.

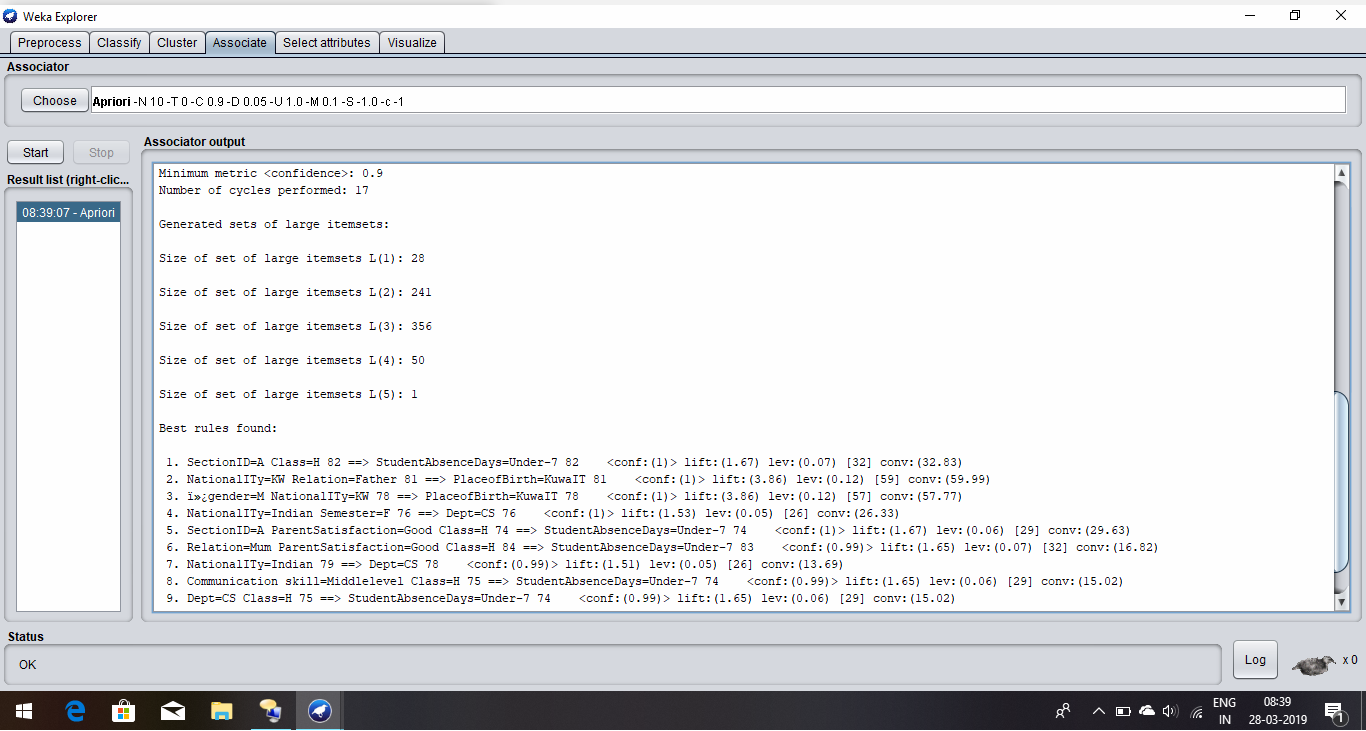
Now if we open the saved result of ‘visualization of classifier errors’ in WEKA ARFF viewer we can see a new attribute named as ‘Predicted RESULT’ is generated in test dataset which maybe or may not be similar with the original result attribute of trained dataset. This known as prediction of WEKA where WEKA predicts the result of student performance which further can be studied for analysis purpose.



**Fig.5.4.2.Plot diagram**

**RESULT OF APRIORI ALGORITHM**:

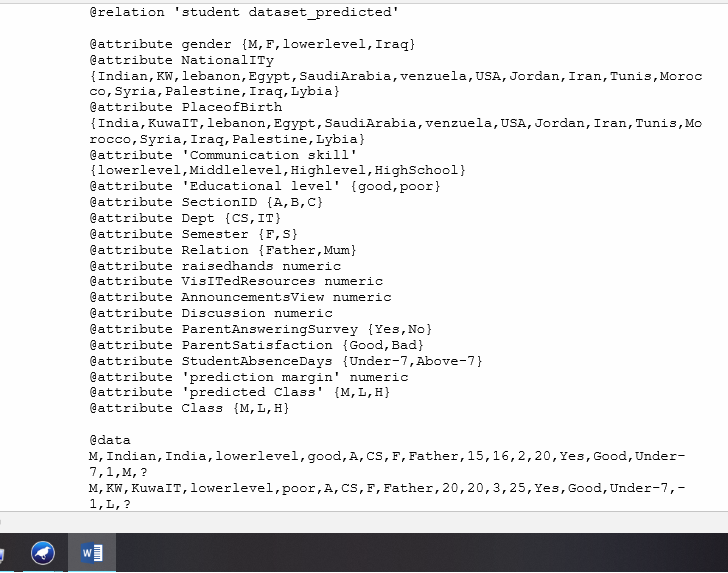




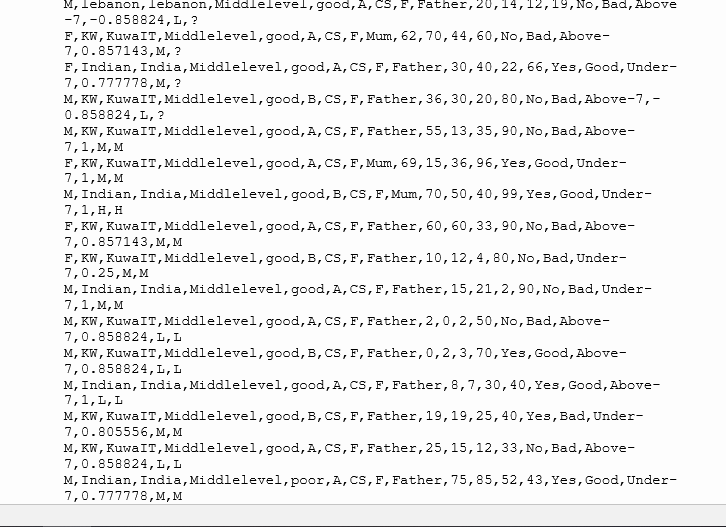
**5.4.3.Apriori Algorithm**

**5.4 Prediction of result**

First, the file with cases to predict needs to have the same structure that the file used to learn the model. The difference is that the value of the result attribute is “?” for all instances (question marks represent missing values in WEKA).

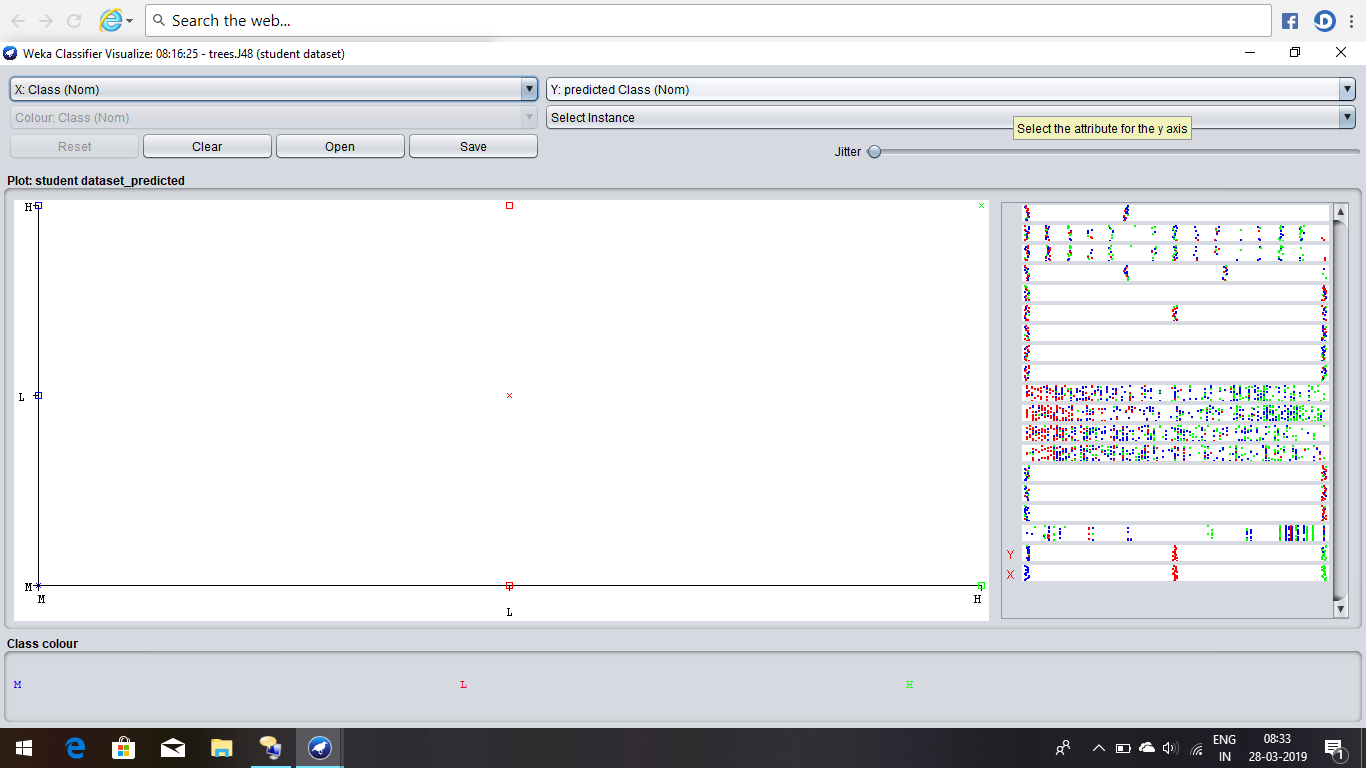


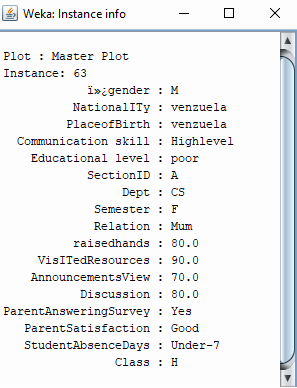
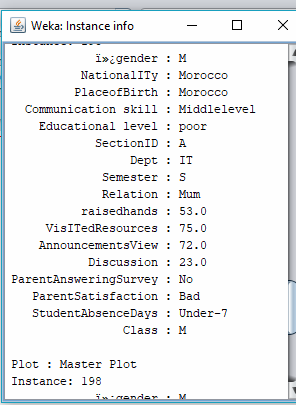
**Fig.5.5.a.Predicted result**



**Fig.5.5.b.Predicted result**

**5.5 VISUALIZATION**



**CHAPTER 6**

**CONCLUSION AND FUTURE ENHANCEMENT**

**6.1 Conclusion**

Educational data mining main focus is to analyse the education system. This paper demonstrates classification method to predict student performance. It also assist in automating the existing manual system by providing the Web Based Information System. It creates connectivity between Alumni and college. All the stakeholders, faculty and management can get the required information without delay. Thus improving their standards and performance.The results of the data mining algorithms for the classification of the students based on the attributes selected reveals that the prediction rates are not uniform. It analysing records of students extra- curricular skills and provide a suggestions on communication and technical skill development by which students can be built in professional aspect of talents.

**6.2 Future Enhancement**

The work can be further extended out by web based application. In future work the authors also interested in working in future on data of students assessments for each course trying to know what kind of student succeed on what kind of courses. It may define what kinds of courses are adapted for every student’s model who shares the same characteristics. It may also provide various multidimensional summary reports and redefine pedagogical learning paths.

**APPENDICES**

**SOURCE CODE:**

**LOGIN**

import static DbConnect.DB.Con;

import java.io.IOException;

import java.io.PrintWriter;

import java.sql.Connection;

import java.sql.PreparedStatement;

import java.sql.ResultSet;

import java.sql.SQLException;

import java.util.logging.Leve

l;

import java.util.logging.Logger;

import javax.servlet.RequestDispatcher;

import javax.servlet.ServletException;

import javax.servlet.http.HttpServlet;

import javax.servlet.http.HttpServletRequest;

import javax.servlet.http.HttpServletResponse;

import javax.servlet.http.HttpSession;

public class Login extends HttpServlet {

String uname,pass,type;

protected void processRequest(HttpServletRequest request, HttpServletResponse response)

throws ServletException, IOException {

response.setContentType("text/html;charset=UTF-8");

PrintWriter out = response.getWriter();

{

HttpSession session = request.getSession();

uname=request.getParameter("uname");

pass=request.getParameter("pass");

type=request.getParameter("type");

Connection con = Con();

if(type.equals("buyer"))

{

try

{

PreparedStatement query=Con().prepareStatement("select \* from ureg where uname='"+uname+"' and pass='"+pass+"' ");

ResultSet rs = query.executeQuery();

if(rs.next())

{

session.setAttribute("uname", uname);

out.println("<script type=\"text/javascript\">");

out.println("alert(\"Welcome "+uname+"\")");

out.println("</script>");

RequestDispatcher rd=request.getRequestDispatcher("buyer.jsp");

rd.include(request, response);

}

else

{

out.println("<script type=\"text/javascript\">");

out.println("alert(\"Please Check Your Username and Password.\")");

out.println("</script>");

RequestDispatcher rd=request.getRequestDispatcher("index.jsp");

rd.include(request, response);

}

} catch (SQLException ex)

{

System.out.println(ex);

}

finally

{

try {

con.close();

} catch (SQLException ex) {

Logger.getLogger(Login.class.getName()).log(Level.SEVERE, null, ex);

}

}

}

else if(type.equals("broker")&&uname.equals("staff")&&pass.equals("staff"))

{

out.println("<script type=\"text/javascript\">");

out.println("alert(\"Welcome Staff\")");

out.println("</script>");

RequestDispatcher rd=request.getRequestDispatcher("broker.jsp");

rd.include(request, response);

}

else

{

out.println("<script type=\"text/javascript\">");

out.println("alert(\"Invalid Credentials\")");

out.println("</script>");

RequestDispatcher rd=request.getRequestDispatcher("index.jsp");

rd.include(request, response);

}

}

}

protected void doGet(HttpServletRequest request, HttpServletResponse response)

throws ServletException, IOException {

processRequest(request, response);

}

protected void doPost(HttpServletRequest request, HttpServletResponse response)

throws ServletException, IOException {

processRequest(request, response);

}

public String getServletInfo() {

return "Short description";

}

}

**REGISTER**

import static DbConnect.DB.Con;

import java.io.IOException;

import java.io.PrintWriter;

import java.sql.Connection;

import java.sql.PreparedStatement;

import java.sql.SQLException;

import java.util.logging.Level;

import java.util.logging.Logger;

import javax.servlet.RequestDispatcher;

import javax.servlet.ServletException;

import javax.servlet.http.HttpServlet;

import javax.servlet.http.HttpServletRequest;

import javax.servlet.http.HttpServletResponse;

public class Register extends HttpServlet {

String uname,pass,email,phone,location,state,type;

protected void processRequest(HttpServletRequest request, HttpServletResponse response)

throws ServletException, IOException {

response.setContentType("text/html;charset=UTF-8");

PrintWriter out = response.getWriter();

{

uname=request.getParameter("uname");

pass=request.getParameter("pass");

email=request.getParameter("email");

phone=request.getParameter("phone");

location=request.getParameter("location");

state=request.getParameter("state");

type=request.getParameter("type");

Connection con = Con();

if(type.equals("buyer"))

{

try

{

PreparedStatement query=con.prepareStatement("insert into ureg (uname,pass,email,phone,location,state) values ('"+uname+"','"+pass+"','"+email+"','"+phone+"','"+location+"','"+state+"') ");

query.executeUpdate();

out.println("<script type=\"text/javascript\">");

out.println("alert(\"Registered Successfully\")");

out.println("</script>");

RequestDispatcher rd=request.getRequestDispatcher("slogin.jsp");

rd.include(request, response);

} catch (SQLException ex)

{

System.out.println(ex);

}

finally

{

try {

con.close();

} catch (SQLException ex) {

Logger.getLogger(Register.class.getName()).log(Level.SEVERE, null, ex);

}

}

}

}

}

@Override

protected void doGet(HttpServletRequest request, HttpServletResponse response)

throws ServletException, IOException {

processRequest(request, response);

}

protected void doPost(HttpServletRequest request, HttpServletResponse response)

throws ServletException, IOException {

processRequest(request, response);

}

@Override

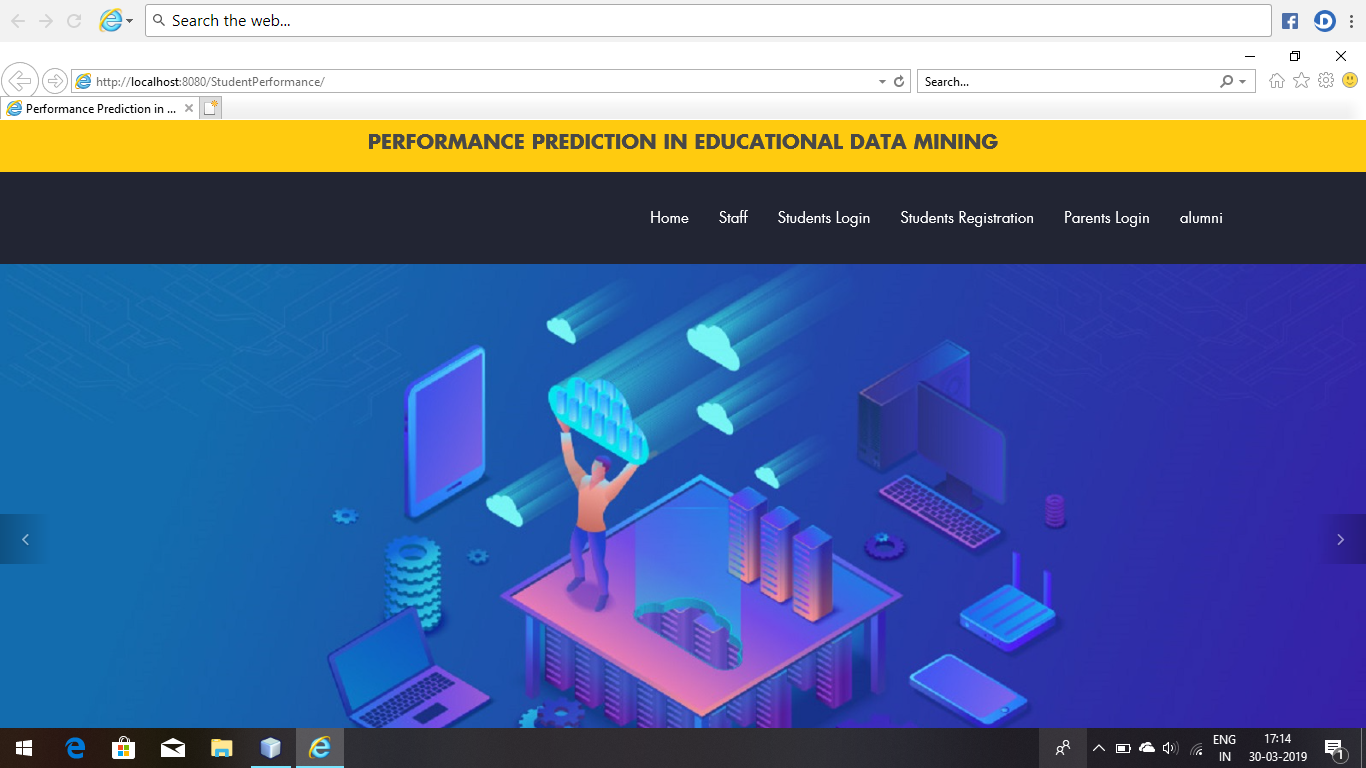
public String getServletInfo() {

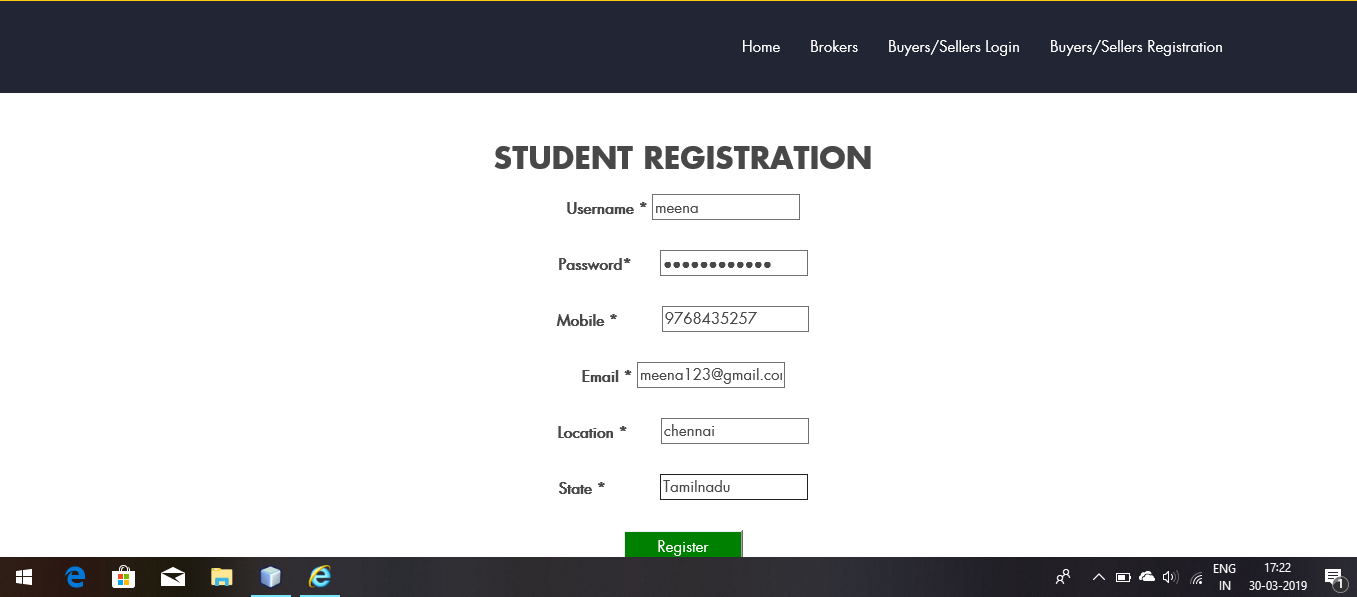
return "Short description";

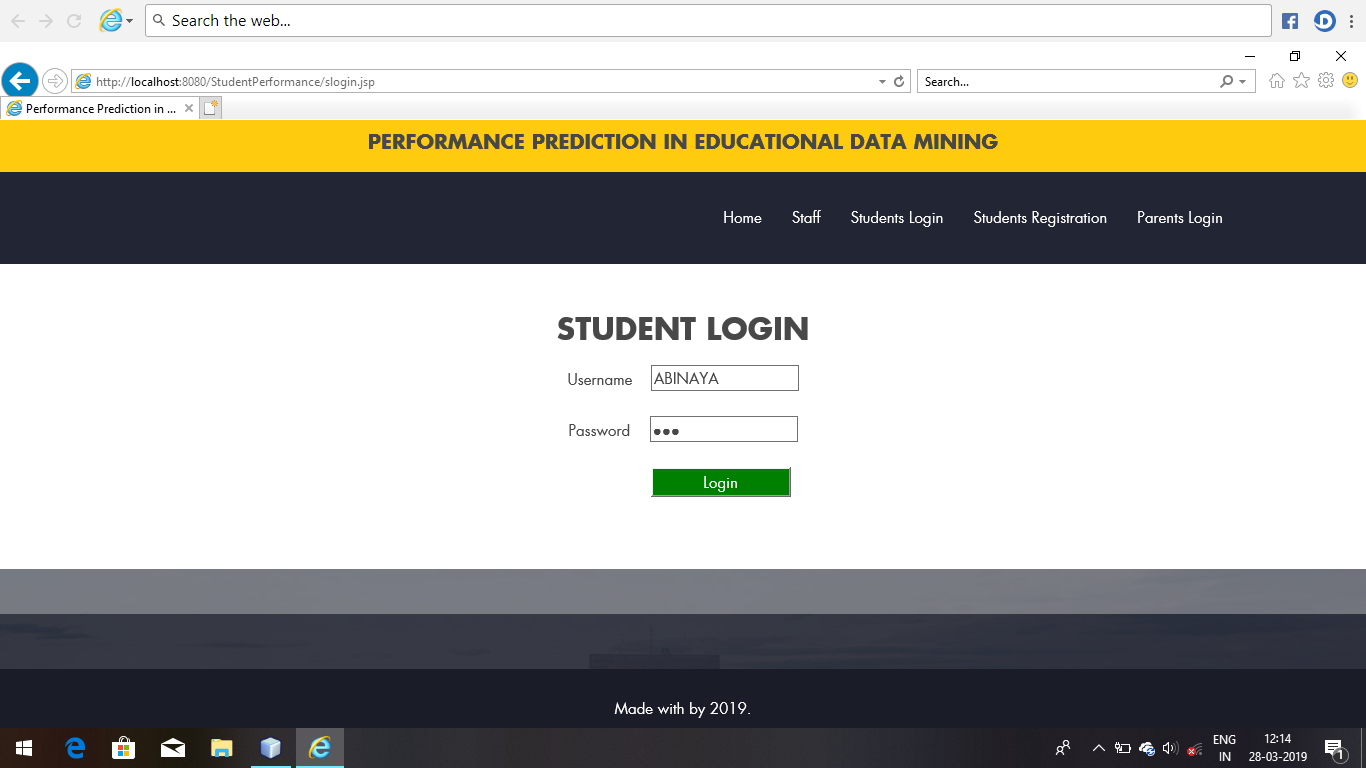
}

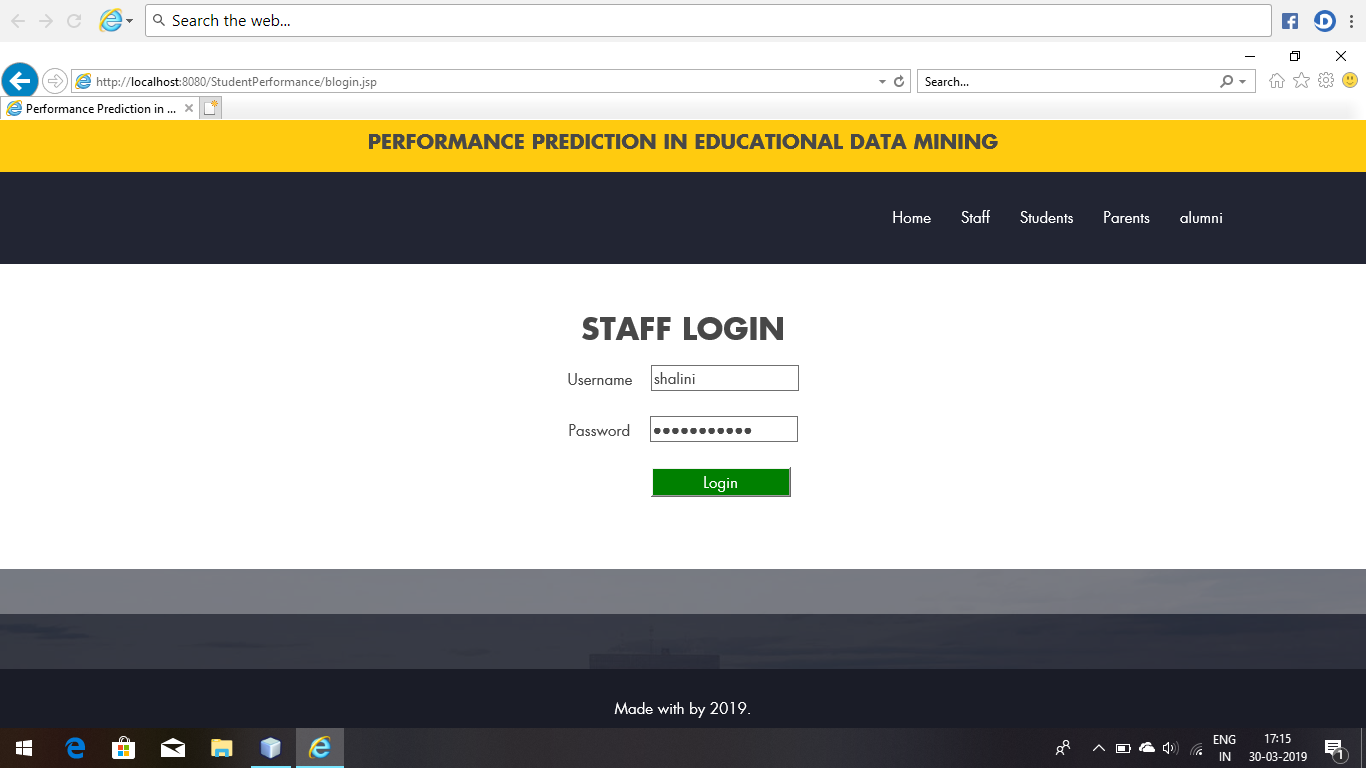
}

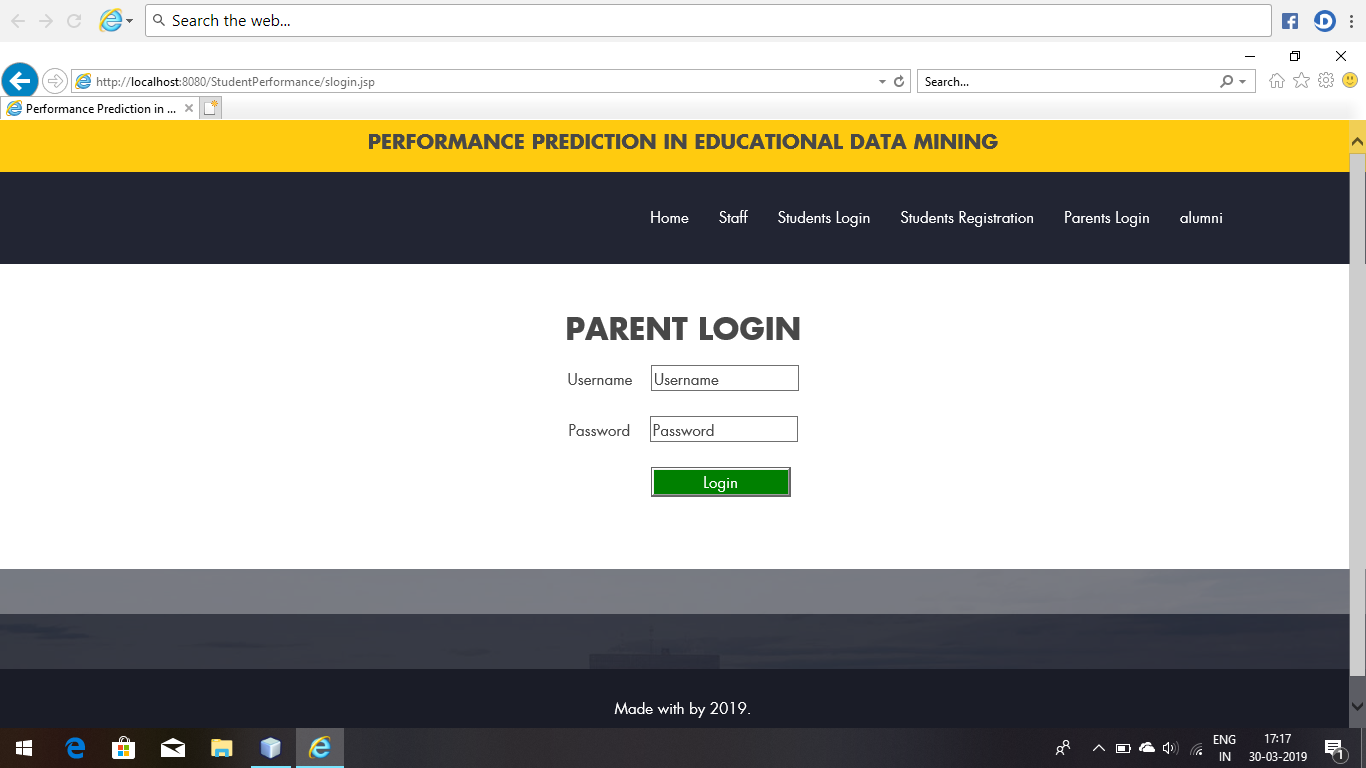
**SCREENSHOTS**

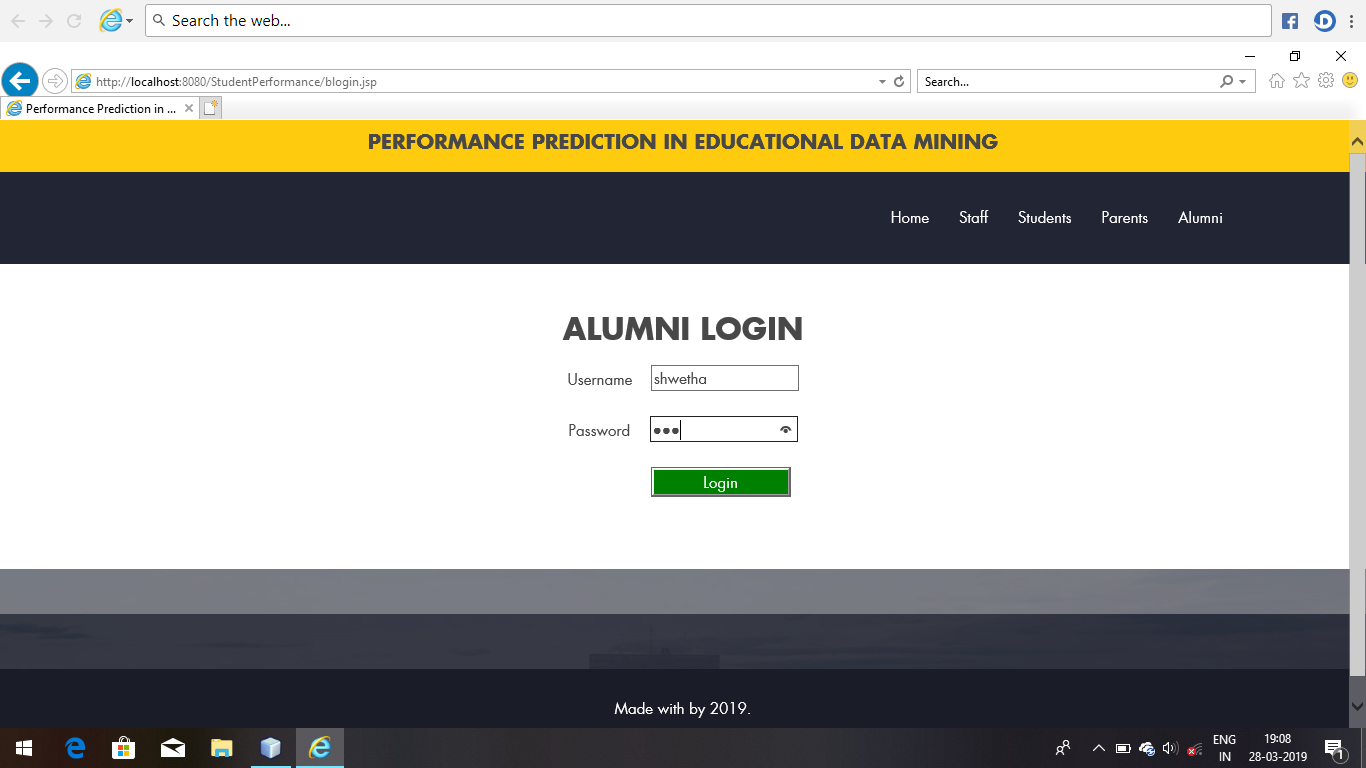


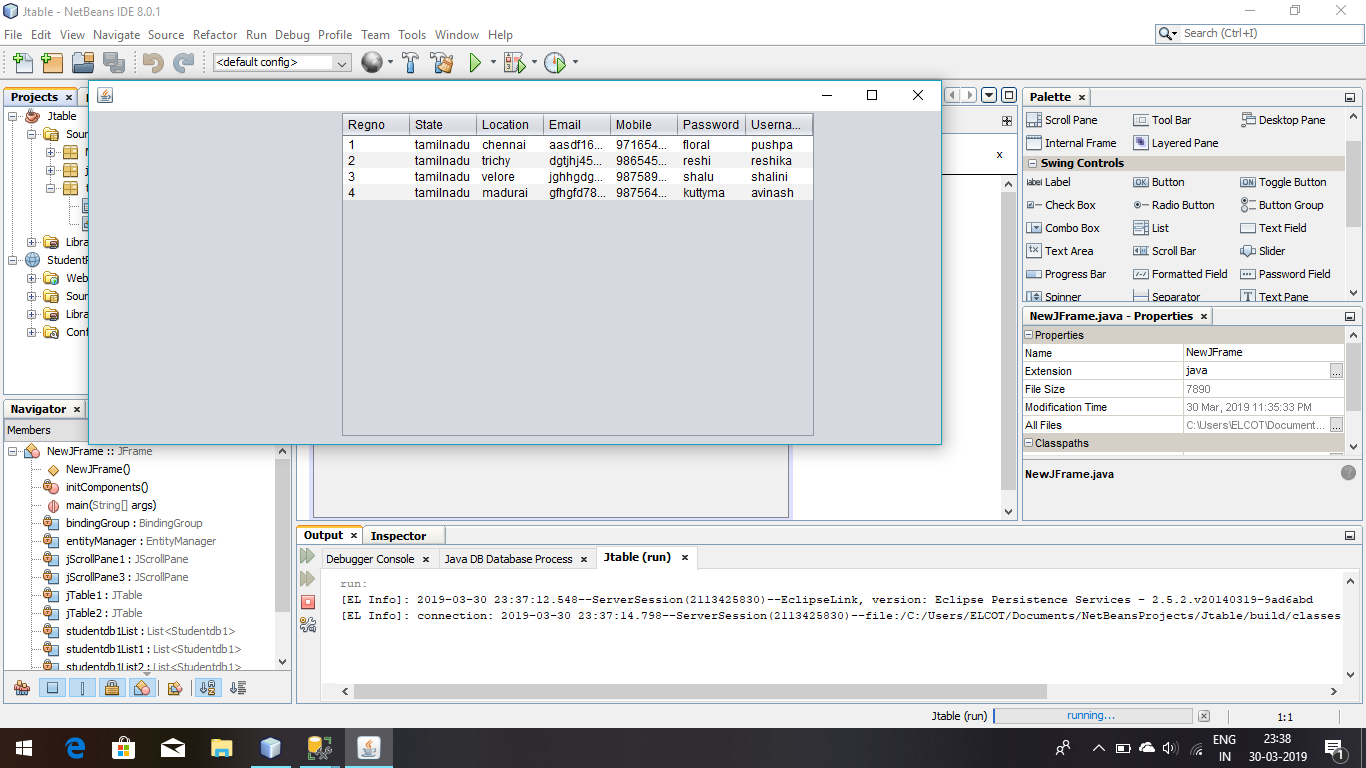












**REFERENCES:**

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

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

[3] Piatetsky-Shapiro, Gregory (1991), Discovery, analysis, and presentation of strong rules, in Piatetsky-Shapiro, Gregory; and Frawley, William J.; eds., Knowledge Discovery in Databases, AAAI/MIT Press, Cambridge, MA

[4] Bo Guo, Rui Zhang, Guang Xu, Chuangming Shi, Li Yang, “Predicting Students Performance in Educational Data Mining”, 2015 International Symposium on Educational Technology , 978-1- 4673-7370-8/15©2015 IEEE.

[5] Krina Parmar Prof. Dineshkumar Vaghela Dr Priyanka Sharma, “Performance Prediction of students using Distributed Data mining”, IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems, 978-1-4799-6818-3/15 © 2015 IEEE

[6] C. Anuradha, T. Velmurugan “A Data Mining based Survey on Student Performance Evaluation System.” 2014 IEEE International Conference on Computational Intelligence and Computing Research, 978-1-4799-3975-6/14 ©2014 IEEE

[7] R. Sumitha, E.S. Vinothkumar “Prediction of Students Outcome Using Data Mining Techniques”International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue- 6,June 2016

[8] Zhibing Liu, Huixia Wang,Hui Zan “Design and implementation of student information management system.” 2010 International symposium on intelligence information processing and trusted computing. 978-0-7695- 4196-9/10 IEEE

[9] S.R.Bharamagoudar, Geeta R.B., S.G.Totad, “Web Based Student Information Management System”,International Journal of Advanced Research in Computer and Communication EngineeringVol.2,Issue6,June2013

[10] Manasi Kawathekar, Kirti K. Bhate and Pankaj Belgoankar “An Android Application for Student Information System” International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 9, Sept 2015

[11] R. R. Kabra And R. S. Bichkar(2011),‖Performance Prediction Of Engineering Students Using Decision Trees‖, International Journal Of Computer Applications (0975 – 8887) Volume 36– No.11, December 2011.