

A REVIEW ON DATA MINING AND MACHINE LEARNING METHODS FOR STUDENT SCHOLARSHIP PREDICTION

Submitted in partial fulfillment of the requirements for
the award of
Bachelor of Engineering degree in Computer Science and Engineering

by

Yadla Venkatesh (38110657)
Pachigolla Rakesh (38110376)



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A" by NAAC

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI - 600 119**

MARCH - 2022



SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of Yadla Venkatesh (38110657) , Pachigolla Rakesh (38110376) who carried out the project entitled **“A REVIEW ON DATA MINING AND MACHINE LEARNING METHODS FOR STUDENT SCHOLARSHIP PREDICTION”** under my supervision from October 2021 to May 2022.

Internal Guide

Head of the Department

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I **Yadla Venkatesh , Pachigolla Rakesh** hereby declare that the Project Report entitled **A REVIEW ON DATA MINING AND MACHINE LEARNING METHODS FOR STUDENT SCHOLARSHIP PREDICTION** done by me under the guidance of **Dr.M.Saravanan M.E.,Ph.D.,**(Internal) is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering / Technology degree in **COMPUTER SCIENCE AND ENGINEERING.**

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph.D**, Dean, School of Computing **Dr. L. Lakshmanan M.E., Ph.D. ,** and **Dr.S.Vigneshwari M.E., Ph.D.** Heads of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. M.Saravanan M.E.,Ph.D.**, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

Most of the schemes introduced by government go into the dustbin just because the officials who implement the schemes could not make them available to suitable people. So there is a secured and transparent system needed which enable an arbitrary person to directly apply for a scheme and track the status from time to time and know whether he is entitled to receive the fruit or his application is rejected by officials. In our system admin will add the scheme details to the system. And he can able to the view the registered user details. And he can able to accept or reject the scheme which is requested by the client. And user can able to view the status of his scheme request level. It contains a small description of ML/DM which are used by the researchers. It also describes data sets as very important in ML/DM methods. Machine Learning becomes most popular in the field of IT industry. Nowadays Machine Learning and Data Mining turn as a powerful technique which applicable for various fields such as IT, Education sector and also in business sector too. The different types of ML/DM algorithms are addressed by using all this technique. The algorithms which give more accuracy results in detection of continuity of every student's scholarship such as Naïve Bayes, Decision Tree and k-NN. Finally, the proposed model will provide a list of candidates, who deserve to have a scholarship and also discussion has been made on accuracy of each techniques which was used to get a result.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	i
	LIST OF FIGURES	ii
1	INTRODUCTION	1
	1.1 SCOPE OF THE PROJECT	1
	1.2 PROPOSED ALGORITHMS	1
2	LITERATURE SURVEY	9
3	SYSTEM REQUIREMENTS	12
	3.1 HARDWARE REQUIREMENTS	12
	3.2 SOFTWARE REQUIREMENTS	12
	3.3 SOFTWARE ENVIRONMENT	12
	3.4 FEATURES OF JAVA	13
	3.5 COLLECTION OF FRAMEWORK	15
	3.6 MY SQL	16
	3.6.1 SQL SERVER MANAGEMENT STUDIO	17
	3.6.2 CREATE A NEW DATABASE	18
	3.7 NETBEANS	20
	3.8 FEASIBILITY STUDY	21
	3.8.1 ECONOMICAL FEASIBILITY	22
	3.8.2 TECHNICAL FEASIBILITY	22
	3.8.3 SOCIAL FEASIBILITY	22
4	SYSTEM ANALYSIS	23
	4.1 PURPOSE	23
	4.2 SCOPE	23
	4.3 EXISTING SYSTEM	23

	4.4 PROPOSED SYSTEM	23
5	SYSTEM DESIGN	25
	5.1 INPUT DESIGN	25
	5.2 OUTPUT DESIGN	25
	5.3 DATA FLOW DIAGRAM	26
	5.4 UML DIAGRAMS	26
	5.4.1 USE CASE DIAGRAM	27
	5.4.2 ACTIVITY DIAGRAM	28
	5.4.3 CLASS DIAGRAM	32
6	MODULES	33
7	SYSTEM IMPLEMENTATION	35
8	SYSTEM TESTING	36
	8.1 TEST PLAN	36
	8.2 VERIFICATION	36
	8.4 BASICS OF SOFTWARE TESTING	36
	8.5 BLACK BOX TESTING	36
	8.6 WHITE BOX TESTING	37
	8.7 TYPES OF TESTING	37
	8.8 REQUIREMENT ANALYSIS	39
	8.8.1 FUNCTIONAL REQUIREMENTS	39
	8.8.2 NON- FUNCTIONAL REQUIREMENTS	40
9	CONCLUSION	42
	REFERENCES	43
	APPENDICES	
	SAMPLE CODE	44
	SCREEN SHOTS	51
	PUBLICATION WITH PLAGIARISM REPORT	

LIST OF FIGURES

FIGURE No.	FIGURE NAME	PAGE NO
1.2.1	Fig 1.2.1 Plotted dataset	3
1.2.2	Fig 1.2.2 Plotted data point	3
3.4	Fig 3.4 Java Compilation	14
3.5	Fig 3.5 Collection Framework	15
3.6	Fig 3.6 SQL Server Management	17
3.6.1	Fig 3.6.1 Query Editor	18
3.6.2	Fig 3.6.2 Creating a New Database	19
5.4.1	Fig 5.4.1 Use Case Diagram	28
5.4.2	Fig 5.4.2 Activity Diagrams	29-31
5.4.3	Fig 5.4.3 Class Diagrams	32
7	Fig.7 System Architecture	35

CHAPTER 1

INTRODUCTION

The objective of the system is to develop a system which provides the details of government schemes and beneficiaries. The purpose is to design a system which automates the processes involved and allows users to perform various operations. The purpose is to design a system which automates the processes involved and allows users to perform various operations.

1.1 SCOPE OF THE PROJECT

The scope of Government Schemes Alerts is as follows: The Government Schemes Alerts provides an easier and save the time to register and use the scheme services. The scope of the government scheme alerts is to assure a transparent and secured system. The Government Schemes Alerts helps user to directly apply for a scheme and track the status from time to time. Through this site helps to know whether user is entitled to receive the fruit or his application is rejected by officials. This site is helpful to every individual to check their current status of their scheme request level.

1.2 PROPOSED ALGORITHMS

KNN

K-nearest neighbors (KNN) algorithm is a type of supervised ML algorithm which can be used for both classification as well as regression predictive problems. However, it is mainly used for classification predictive problems in industry. The following two properties would define KNN well –

- Lazy learning algorithm – KNN is a lazy learning algorithm because it does not have a specialized training phase and uses all the data for training while classification.
- Non-parametric learning algorithm – KNN is also a non-parametric learning algorithm because it doesn't assume anything about the underlying data.

Working of KNN Algorithm:

K-nearest neighbors (KNN) algorithm uses 'feature similarity' to predict the values of new datapoints which further means that the new data point will be assigned a value based on how closely it matches the points in the training set. We can understand its working with the help of following steps –

Step 1 – For implementing any algorithm, we need dataset. So during the first step of KNN, we must load the training as well as test data.

Step 2 – Next, we need to choose the value of K i.e. the nearest data points. K can be any integer.

Step 3 – For each point in the test data do the following –

- **3.1** – Calculate the distance between test data and each row of training data with the help of any of the method namely: Euclidean, Manhattan or Hamming distance. The most commonly used method to calculate distance is Euclidean.
- **3.2** – Now, based on the distance value, sort them in ascending order.
- **3.3** – Next, it will choose the top K rows from the sorted array.
- **3.4** – Now, it will assign a class to the test point based on most frequent class of these rows.

Step 4 – End

Example

The following is an example to understand the concept of K and working of KNN algorithm –

Suppose we have a dataset which can be plotted as follows –

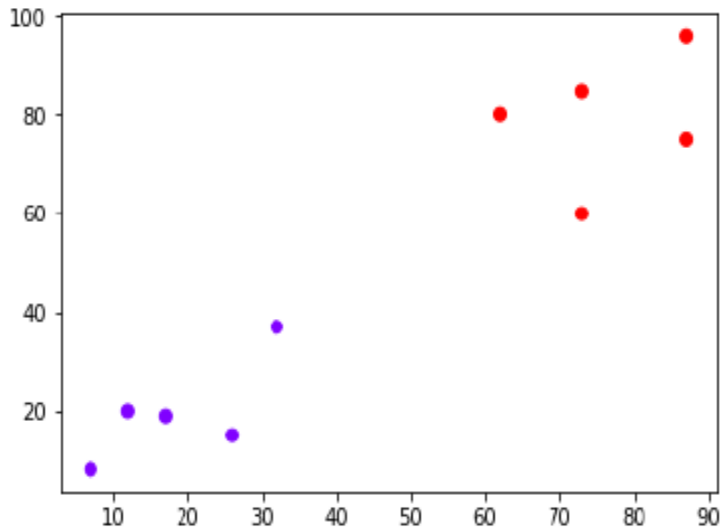


Fig 1.2.1 Plotted dataset

Now, we need to classify new data point with black dot (at point 60,60) into blue or red class. We are assuming $K = 3$ i.e. it would find three nearest data points. It is shown in the next diagram –

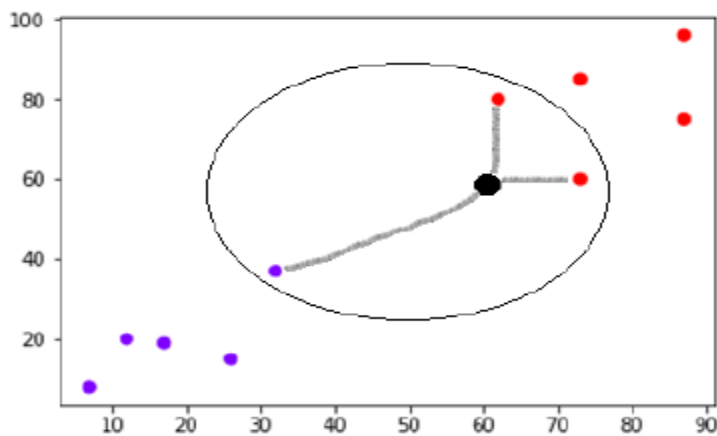


Fig 1.2.2 Plotted data point

We can see in the above diagram the three nearest neighbours of the data point with black dot. Among those three, two of them lies in Red class hence the black dot will also be assigned in red class.

Implementation in Python

As we know K-nearest neighbors (KNN) algorithm can be used for both classification as well as regression. The following are the recipes in Python to use KNN as classifier as well as regressor –

KNN as Classifier

First, start with importing necessary python packages –

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Next, download the iris dataset from its weblink as follows –

```
path = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
```

Next, we need to assign column names to the dataset as follows –

```
headers = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'Class']
```

Now, we need to read dataset to pandas dataframe as follows –

```
dataset = pd.read_csv(path, names = headers)
dataset.head()
```

	sepal-length	sepal-width	petal-length	petal-width	Class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

Data Preprocessing will be done with the help of following script lines.

```
X = dataset.iloc[:, :-1].values
```

```
y = dataset.iloc[:, 4].values
```

Next, we will divide the data into train and test split. Following code will split the dataset into 60% training data and 40% of testing data –

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.40)
```

Next, data scaling will be done as follows –

```
from sklearn.preprocessing import StandardScaler
```

```
scaler = StandardScaler()
```

```
scaler.fit(X_train)
```

```
X_train = scaler.transform(X_train)
```

```
X_test = scaler.transform(X_test)
```

Next, train the model with the help of KNeighborsClassifier class of sklearn as follows –

```
from sklearn.neighbors import KNeighborsClassifier
```

```
classifier = KNeighborsClassifier(n_neighbors = 8)
```

```
classifier.fit(X_train, y_train)
```

At last we need to make prediction. It can be done with the help of following script –

```
y_pred = classifier.predict(X_test)
```

Next, print the results as follows –

```
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
```

```
result = confusion_matrix(y_test, y_pred)
```

```
print("Confusion Matrix:")
```

```
print(result)
```

```
result1 = classification_report(y_test, y_pred)
```

```
print("Classification Report:",)
```

```
print (result1)
```

```
result2 = accuracy_score(y_test,y_pred)
```

```
print("Accuracy:",result2)
```

Output

Confusion Matrix:

```
[[21 0 0]
 [ 0 16 0]
 [ 0 7 16]]
```

Classification Report:

	precision	recall	f1-score	support
Iris-setosa	1.00	1.00	1.00	21
Iris-versicolor	0.70	1.00	0.82	16
Iris-virginica	1.00	0.70	0.82	23
micro avg	0.88	0.88	0.88	60
macro avg	0.90	0.90	0.88	60
weighted avg	0.92	0.88	0.88	60

Accuracy: 0.8833333333333333

KNN as Regressor

First, start with importing necessary Python packages –

```
import numpy as np
```

```
import pandas as pd
```

Next, download the iris dataset from its weblink as follows –

```
path = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
```

Next, we need to assign column names to the dataset as follows –

```
headernames = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'Class']
```

Now, we need to read dataset to pandas dataframe as follows –

Next, import *KNeighborsRegressor* from *sklearn* to fit the model –

```
from sklearn.neighbors import KNeighborsRegressor
```

```
knnr = KNeighborsRegressor(n_neighbors = 10)
knnr.fit(X, y)
```

At last, we can find the MSE as follows –

```
print ("The MSE is:",format(np.power(y-knnr.predict(X),2).mean()))
```

Output

The MSE is: 0.12226666666666669

Pros and Cons of KNN

Pros

- It is very simple algorithm to understand and interpret.
- It is very useful for nonlinear data because there is no assumption about data in this algorithm.
- It is a versatile algorithm as we can use it for classification as well as regression.
- It has relatively high accuracy but there are much better supervised learning models than KNN.

Cons

- It is computationally a bit expensive algorithm because it stores all the training data.
- High memory storage required as compared to other supervised learning algorithms.
- Prediction is slow in case of big N.
- It is very sensitive to the scale of data as well as irrelevant features.

Applications of KNN

The following are some of the areas in which KNN can be applied successfully

Banking System

KNN can be used in banking system to predict whether an individual is fit for loan approval? Does that individual have the characteristics similar to the defaulters one?

Calculating Credit Ratings

KNN algorithms can be used to find an individual's credit rating by comparing with the persons having similar traits.

CHAPTER 2

LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system. The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

The Incentive of Government Chief Officials at Grass-roots Level: An Analysis of the Political Tournament Model in Interagency Information Sharing

The lack of incentive mechanism for government chief officials is the key obstacle to improving interagency government information sharing. This paper considers the political tournament model as suitable to stimulate local governmental “top leaders” at district-level and county-level, and then puts forward some measures for the application of the model.

Research on the E-Government Scheme Based on Multi-Technologies and Bi-Directional Authentication

The authentication is the key part of the network security. It is also the main problem we should resolve. This scheme, which is bi-directional, using multi-mechanism to ensure safety authentication and providing a uniform authentication service to other e-government application system, meets the need of current e-government system. But authentication technology and e-government system are continually developing, there are still many problems need to be solved.

Analysis of E-government Services Outsourcing and Incentive Scheme

Relative theories, operation patterns and trends about e-government outsourcing are studied. Subsequently, the analyses on the incentive model and system for government and enterprise in the e-government outsourcing project are presented based on the theories of game and principal-agent, emphasizing on motivation and performance evaluation in managing out-sourcing business, in which the importance of incentive game in principal-agent is also derived.

The state of incomplete and asymmetric information makes it important for government to implement reward and punishment before enough analyzing. From the aspect of validity analysis, the optimal proposal will have the characteristics of monotone, that is, the awards to the manager which are provided by government should increase with the improvement of output level. It is eventually concluded that the reformation of e-government services outsourcing relationship requires incentive scheme.

A Risk Assessment Model Based-Business-Circle of E-Government Information System

With the development of computer network and information technology rapidly, they are serving every walk of life in China. Chinese government is using widely the information and network technology for improving its work. So, Chinese E-Government information system is developing feetyly.

However, in fact, E-Government information system has some its characters and some potential hazards in China, which affects E-Government information system's healthy and favorable development. It is insufficient that some secure

products and tools are only used for the security system architecture of E-Government information system. To build secure and authentic system architecture of E-Government information system, the risk assessment and management is necessary for secure and reliable E-Government information system.

A risk assessment model based-business-circle of E-Government information system is proposed in the article according to the characters of E-Government information system. In the model, some business circles are given according to the importance and features of E-Government information system. Moreover, the risk assessment is done in every business circle. The model introduces the idea of risk assessment, scheme of risk assessment, process of risk assessment and the computational method of risk.

A secure publishing scheme in E-Government

One of the most important functions of e-Government is publication through the Internet. Though digital signature can be exploited to protect the integrity and authenticity of published information, it can not effectively resist malicious replacement in scenarios where the published information consists of different files. A secure publishing scheme based on Merkle hash tree is proposed. The scheme is able to resist the malicious replacement of published information and its cost is modest.

CHAPTER 3

SYSTEM REQUIREMENTS

3.1 HARDWARE REQUIREMENTS:

- System - Windows 7/10
- Speed - 2.4GHZ
- Hard disk - 40GB
- Monitor - 15VGA color
- RAM - 4GB

3.2 SOFTWARE REQUIREMENTS:

- Operating System - Windows 10.
- Coding language - JAVA
- IDE - NETBEANS
- Backend - Mysql

3.3 SOFTWARE ENVIRONMENT

Introduction to Java

Java is one of the world's most important and widely used computer languages, and it has held this distinction for many years. Unlike some other computer languages whose influence has wearied with passage of time, while Java's has grown.

Application of Java

Java is widely used in every corner of world and of human life. Java is not only used in software but is also widely used in designing hardware controlling software components. There are more than 930 million JRE downloads each year and 3 billion mobile phones run java.

Following are some other usage of Java:

- Developing Desktop Applications
- Web Applications like Linkedin.com, Snapdeal.com etc
- Mobile Operating System like Android
- Embedded Systems
- Robotics and games etc.

3.4 FEATURES OF JAVA

The prime reason behind creation of Java was to bring portability and security feature into a computer language. Beside these two major features, there were many other features that played an important role in moulding out the final form of this outstanding language. Those features are;

1) Simple

Java is easy to learn and its syntax is quite simple, clean and easy to understand. The confusing and ambiguous concepts of C++ are either left out in Java or they have been re-implemented in a cleaner way.

2) Object Oriented

In java everything is Object which has some data and behaviour. Java can be easily extended as it is based on Object Model.

3) Robust

Java makes an effort to eliminate error prone codes by emphasizing mainly on compile time error checking and runtime checking. But the main areas which Java improved were Memory Management and mishandled Exceptions by introducing automatic Garbage Collector and Exception Handling.

4) Platform Independent

Unlike other programming languages such as C, C++ etc. which are compiled into platform specific machines. Java is guaranteed to be write-once, run-anywhere language.

On compilation Java program is compiled into byte code. This byte code is platform independent and can be run on any machine, plus this byte code format also provide security. Any machine with Java Runtime Environment can run Java Programs.

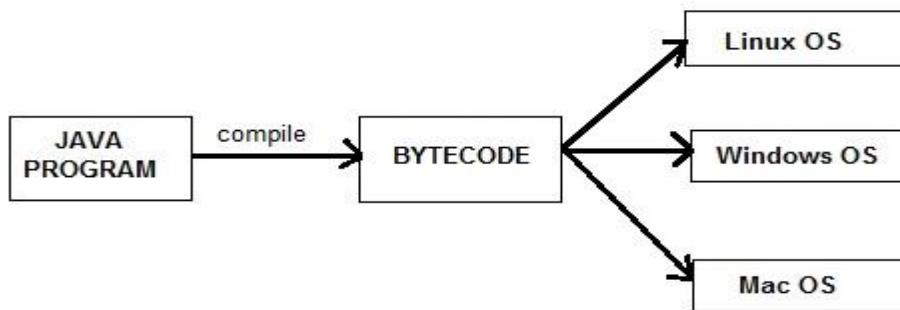


Fig 3.4 Java Compilation

5) Secure

When it comes to security, Java is always the first choice. With java secure features it enable us to develop virus free, temper free system. Java program always runs in Java runtime environment with almost null interaction with system OS, hence it is more secure.

6) Multi-Threading

Java multithreading feature makes it possible to write program that can do many tasks simultaneously. Benefit of multithreading is that it utilizes same memory and other resources to execute multiple threads at the same time, like While typing, grammatical errors are checked along.

7) Architectural Neutral

Compiler generates byte codes, which have nothing to do with a particular computer architecture, hence a Java program is easy to interpret on any machine.

8) Portable

Java Byte code can be carried to any platform. No implementation dependent features. Everything related to storage is predefined, example: size of primitive data types.

10) High Performance

Java is an interpreted language, so it will never be as fast as a compiled language like C or C++. But, Java enables high performance with the use of just-in-time compiler.

3.5 COLLECTION OF FRAMEWORK

Collection framework was not part of original Java release. Collections was added to J2SE 1.2. Prior to Java 2, Java provided adhoc classes such as Dictionary, Vector, Stack and Properties to store and manipulate groups of objects. Collection framework provides many important classes and interfaces to collect and organize group of alike objects.

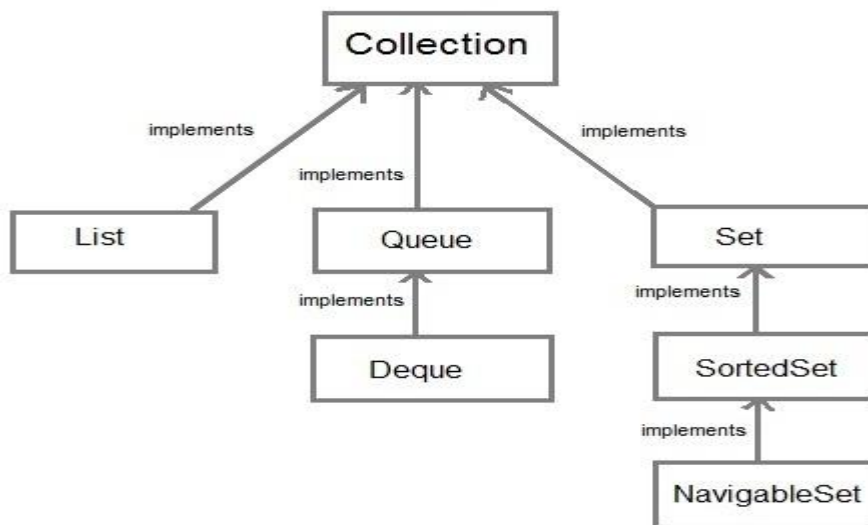


Fig 3.5 Collection Framework

3.6 MY SQL

MySQL, officially, but also called "My Sequel" is the world's most widely used open-source relational database management system (RDBMS) that runs as a server providing multi-user access to a number of databases, though SQLite probably has more total embedded deployments. The SQL phrase stands for Structured Query Language.

The MySQL development project has made its source code available under the terms of the GNU General Public License, as well as under a variety of proprietary agreements. MySQL was owned and sponsored by a single for-profit firm, the Swedish company MySQL AB, now owned by Oracle Corporation. MySQL is a popular choice of database for use in web applications, and is a central component of the widely used LAMP open source web application software stack (and other 'AMP' stacks).

LAMP is an acronym for "Linux, Apache, MySQL and Perl/PHP/Python." Free software-open source projects that require a full-featured database management system often use MySQL.

For commercial use, several paid editions are available, and offer additional functionality. Applications which use MySQL databases include: TYPO3, MODx, Joomla, WordPress, phpBB, MyBB, Drupal and other software. MySQL is also used in many high-profile, large-scale websites, including Wikipedia, Google (though not for searches), Facebook, Twitter, Flickr and YouTube.

T-SQL expands on the SQL standard to include procedural programming, local variables, and various support functions for string processing, date processing, mathematics, etc. SQL Server consists of a Database Engine and a Management Studio (and lots of other stuff which we will not mention here). The Database engine has no graphical interface - it is just a service running in the background of your computer (preferable on the server). The Management Studio is graphical tool for configuring and viewing the information in the database. It can be installed on the server or on the client (or both).

SQL Server Management Studio can also be used to create a new database, alter any existing database schema by adding or modifying tables and indexes, or analyze performance. It includes the query windows which provide a GUI based interface to write and execute queries.

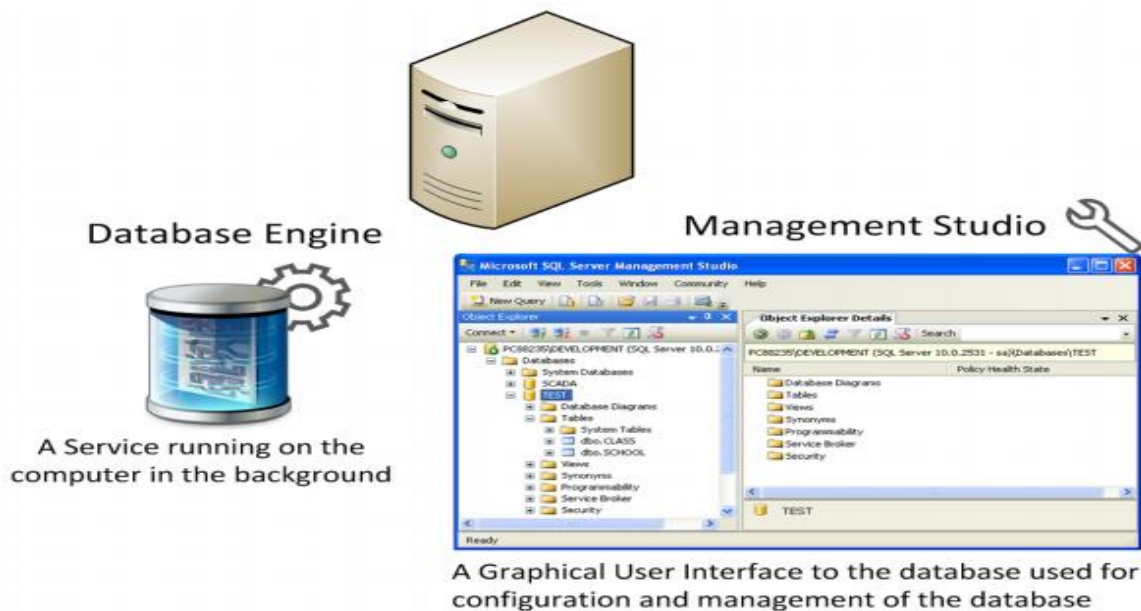


Fig 3.6 SQL Server Management

3.6.1 SQL SERVER MANAGEMENT STUDIO

SQL Server Management Studio is a GUI tool included with SQL Server for configuring, managing, and administering all components within Microsoft SQL Server. The tool includes both script editors and graphical tools that work with objects and features of the server. As mentioned earlier, version of SQL Server Management Studio is also available for SQL Server Express Edition, for which it is known as SQL Server Management Studio Express.

A central feature of SQL Server Management Studio is the Object Explorer, which allows the user to browse, select, and act upon any of the objects within the server. It can be used to visually observe and analyze query plans and optimize the database performance, among others. SQL Server Management Studio can also be used to create a new database, alter any existing database schema by adding or

modifying tables and indexes, or analyze performance. It includes the query windows which provide a GUI based interface to write and execute queries.

When creating SQL commands and queries, the “Query Editor” (select “New Query” from the Toolbar) is used (shown in the figure above). With SQL and the “Query Editor” we can do almost everything with code, but sometimes it is also a good idea to use the different Designer tools in SQL to help us do the work without coding (so much).

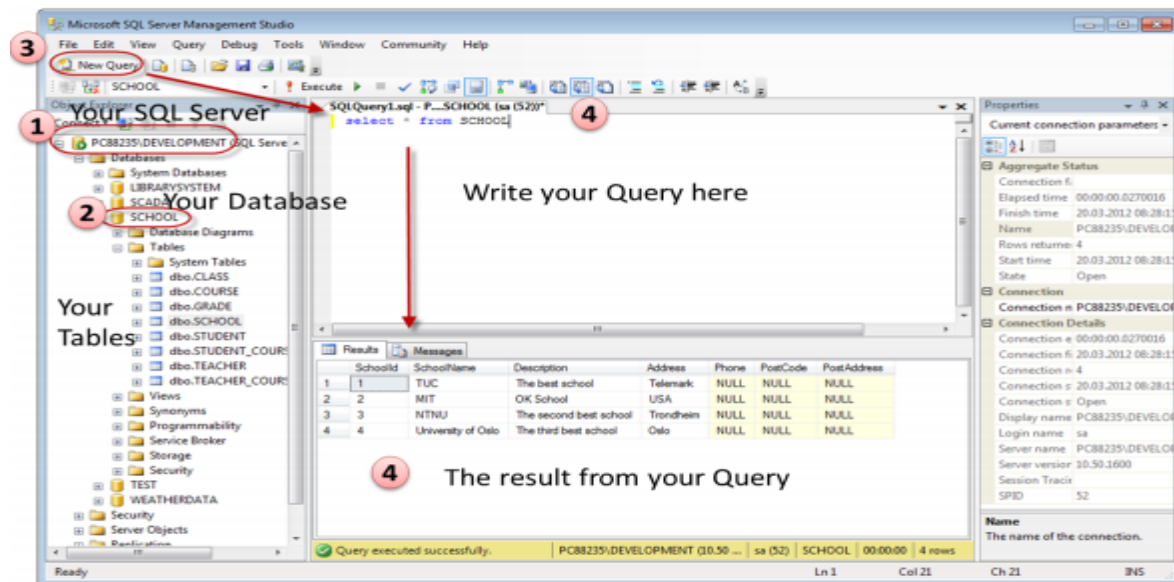


Fig 3.6.1 Query Editor

3.6.2 CREATE A NEW DATABASE

It is quite simple to create a new database in Microsoft SQL Server. Just right-click on the “Databases” node and select “New Database...” The OLAP Services feature available in SQL Server version 7.0 is now called SQL Server 2000 Analysis Services. The term OLAP Services has been replaced with the term Analysis Services. Analysis Services also includes a new data mining component. The Repository component available in SQL Server version 7.0 is now called Microsoft SQL Server 2000 Meta Data Services.

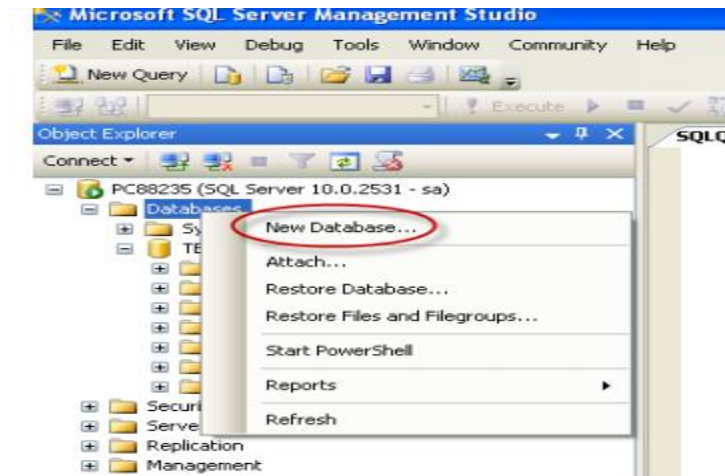


Fig 3.6.2 Creating a New Database

References to the component now use the term Meta Data Services. The term repository is used only in reference to the repository engine within Meta Data Services

SQL-SERVER database consist of 5 type of objects,

They are,

1. TABLE
2. QUERY
3. FORM
4. REPORT
5. MACRO

Table

A database is a collection of data about a specific topic.

Views of Table

We can work with a table in two types,

1. Design View
2. Datasheet View

Design View

To build or modify the structure of a table we work in the table design view. We can specify what kind of data will be hold.

Datasheet View

To add, edit or analyses the data itself we work in tables datasheet view mode.

Query

A query is a question that has to be asked the data. Access gathers data that answers the question from one or more table. The data that make up the answer is either dynaset (if you edit it) or a snapshot (it cannot be edited). Each time we run query, we get latest information in the dynaset. Access either displays the dynaset or snapshot for us to view or perform an action on it, such as deleting or updating.

3.7 NETBEANS

Net Beans IDE is the official IDE for Java 8. With its editors, code analyzers, and converters, you can quickly and smoothly upgrade your applications to use new Java 8 language constructs, such as lambdas, functional operations, and method references. Batch analyzers and converters are provided to search through multiple applications at the same time, matching patterns for conversion to new Java 8 language constructs. With its constantly improving Java Editor, many rich features and an extensive range of tools, templates and samples, Net Beans IDE sets the standard for developing with cutting edge technologies out of the box.

Java Networking

The term network programming refers to writing programs that execute across multiple devices (computers), in which the devices are all connected to each other using a network.

The java.net package provides support for the two common network protocols:

- TCP: TCP stands for Transmission Control Protocol, which allows for reliable communication between two applications. TCP is typically used over the Internet Protocol, which is referred to as TCP/IP.
- UDP: UDP stands for User Datagram Protocol, a connection-less protocol that allows for packets of data to be transmitted between applications.

This tutorial gives good understanding on the following two subjects:

- Socket Programming: This is most widely used concept in Networking and it has been explained in very detail.
- URL Processing: This would be covered separately. [Click here](#) to learn about URL Processing in Java language.

3.8 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.

The feasibility study investigates the problem and the information needs of the stakeholders. It seeks to determine the resources required to provide an information systems solution, the cost and benefits of such a solution, and the feasibility of such a solution.

The goal of the feasibility study is to consider alternative information systems solutions, evaluate their feasibility, and propose the alternative most suitable to the organization. The feasibility of a proposed solution is evaluated in terms of its components.

3.8.1 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 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.

3.8.2 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.

3.8.3 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.

CHAPTER 4

SYSTEM ANALYSIS

4.1 PURPOSE

The purpose of this document is prediction of student scholarship using machine learning and data mining algorithms. In detail, this document will provide a general description of our project, including user requirements, product perspective, and overview of requirements, general constraints. In addition, it will also provide the specific requirements and functionality needed for this project - such as interface, functional requirements and performance requirements.

4.2 SCOPE

The scope of this SRS document persists for the entire life cycle of the project. This document defines the final state of the software requirements agreed upon by the customers and designers. Finally at the end of the project execution all the functionalities may be traceable from the SRS to the product. The document describes the functionality, performance, constraints, interface and reliability for the entire life cycle of the project.

4.3 EXISTING SYSTEM

In the present scenario, various government schemes are published in print media and advertised by the respective departments. It is often difficult for people to get the relevant information and approach the officials for availing the services. Due to lack of proper communication among various departments and officials, often the purpose of the schemes is not achieved. People do not get to know about the schemes and do not benefit from them.

4.4 PROPOSED SYSTEM

The proposed system is a web based application, which maintains a centralized repository of all scheme related information. This makes it easy to search for the desired information. Users and public can access the system to get the information

and also apply online for the proposed schemes. The system allows coordination among all concerned departments and officials. The system also allows the regular monitoring and modification of information.

ADVANTAGES OF PROPOSED SYSTEM

- Provides all the schemes released by the Government.
- Announcing the every new schemes.
- Displaying the notifications on the main pages.
- User registration process.
- Login authentication.
- Allows users to check their status details.

CHAPTER 5

SYSTEM DESIGN

5.1 INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

5.2 OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improve user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the Future.
- Signal important events, opportunities, problems, or warnings.

- Trigger an action.
- Confirm an action

5.3 DATA FLOW DIAGRAM

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

5.4 UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices.

5.4.1 USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

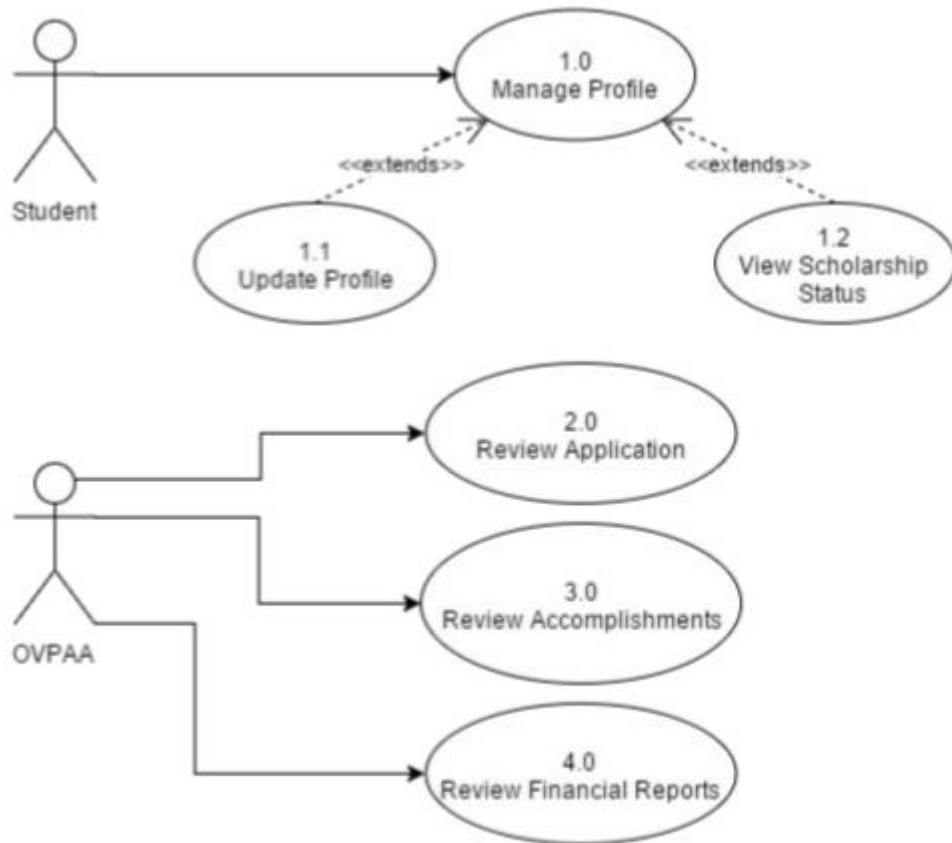
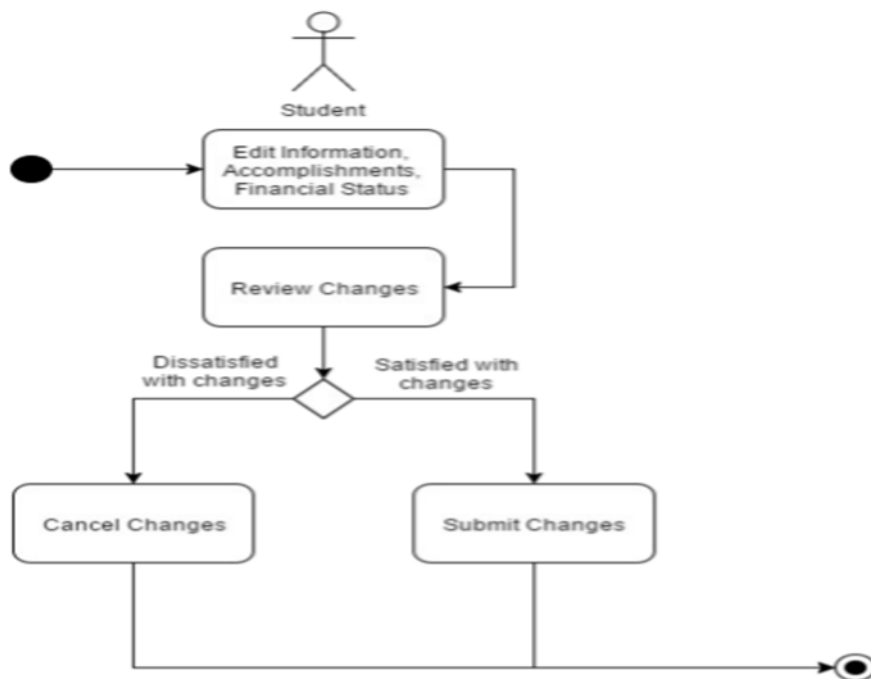
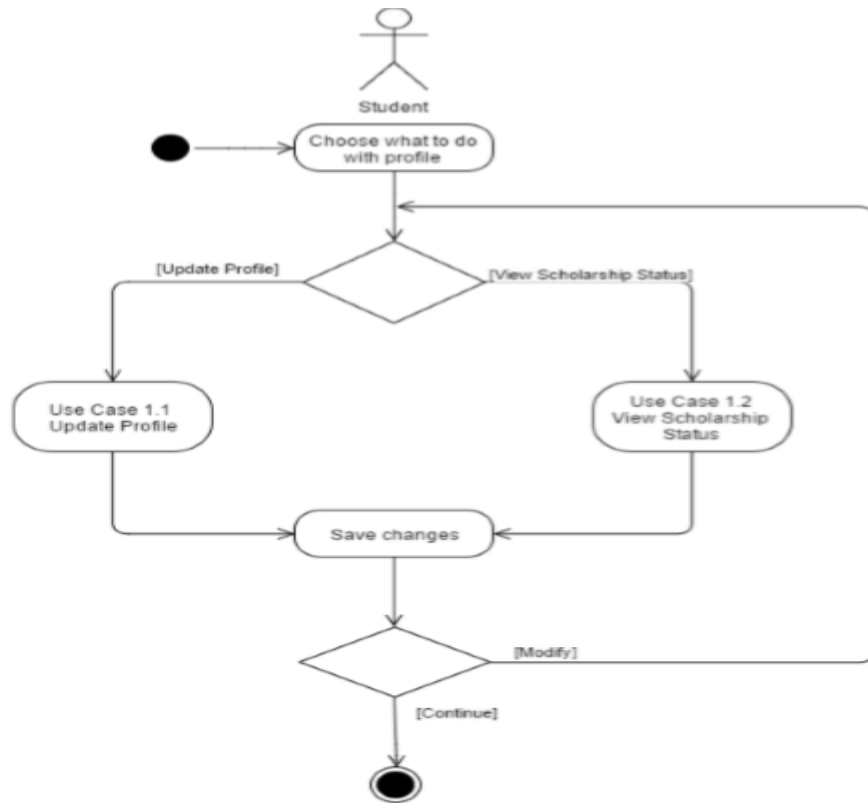
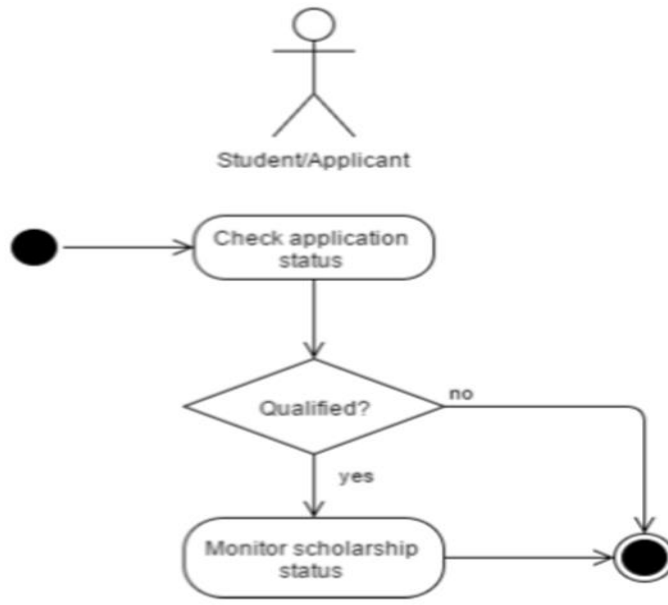


Fig 5.4.1 Use Case Diagram

5.4.2 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.





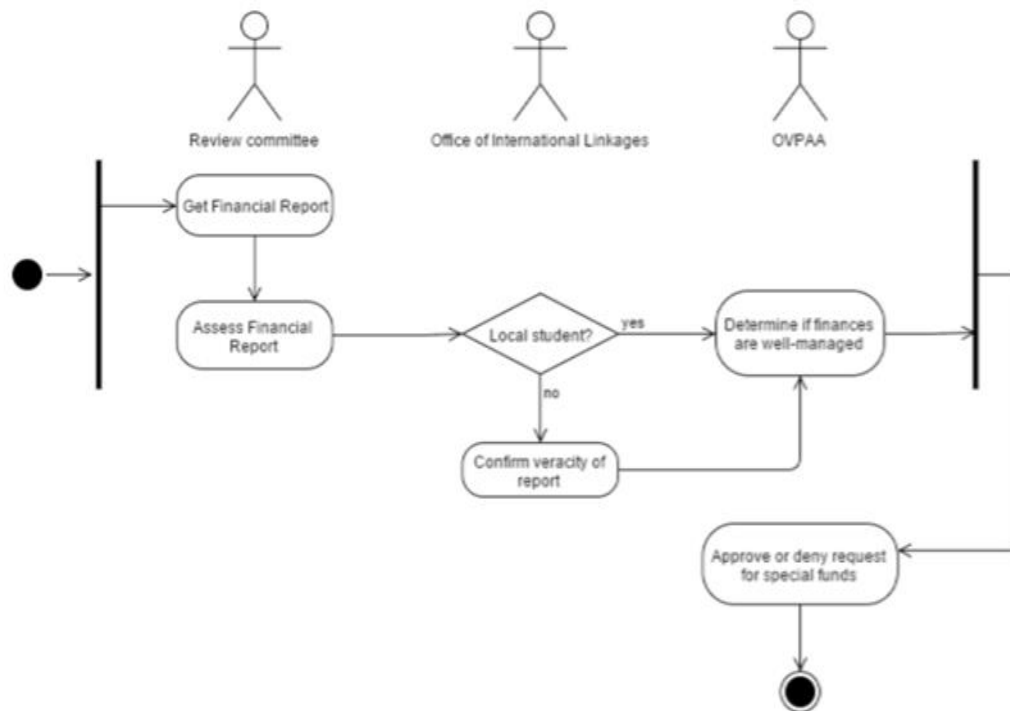


Fig 5.4.2 Activity Diagrams

5.4.3 CLASS DIAGRAM



Fig 5.4.3 Class Diagrams

CHAPTER 6

MODULES

MODULES

- Administrator
- User
- Schemes
- Report
- Result

MODULE DESCRIPTION

Administrator Module

Administrator will check for the new updates are available then admin updates all these information on the site and if user is applicable for the particular schemes administrator send the SMS. Administrator will also check Eligibility details, how many members are already applied for the schemes all these details will be displayed. Admin takes care of the activities related to the schemes provided by the government. This module takes care of accepting various applications from users.

User Module

User got the message if he is illegible for scheme first user visit the website and then checks the document that are needed for schemes and then can apply for that particular scheme. They also check their application status by logging into the system easily. Only user can update or delete the details of users from list.

Schemes Module

This module provides facility to admin, admin can add new scheme, delete previous scheme, update previous scheme. In these is displays al details of scheme also scheme wise user and criteria as well as conditions. Only the admin is able to new scheme on the portal. All the remaining users can view the details of schemes.

Report Module

The reporting module we can see various types of reports like scheme wise report, and user wise report. Various reports like Approved applications, Pending applications and rejected applications also user wise. Scheme wise report contains scheme name, scheme type, scheme id, validity date. User wise report should contains all the information of user and also scheme details which are they apply with criteria and condition. Status of the application and any remarks information will be displayed in approved applications report. All the rejected applications report will be placed in the rejected applications report.

Result Module

To provide that schemes to applicable user first we required that users all information to check every condition mention above, for that we collect users information using user registration process after that we are able to deliver the particular scheme to that particular applicable person or user. Using it user can get benefit from tat government scheme.

CHAPTER 7

SYSTEM IMPLEMENTATION

SYSTEM ARCHITECTURE

Describing the overall features of the software is concerned with defining the requirements and establishing the high level of the system. During architectural design, the various web pages and their interconnections are identified and designed. The major software components are identified and decomposed into processing modules and conceptual data structures and the interconnections among the modules are identified. The following modules are identified in the proposed system.

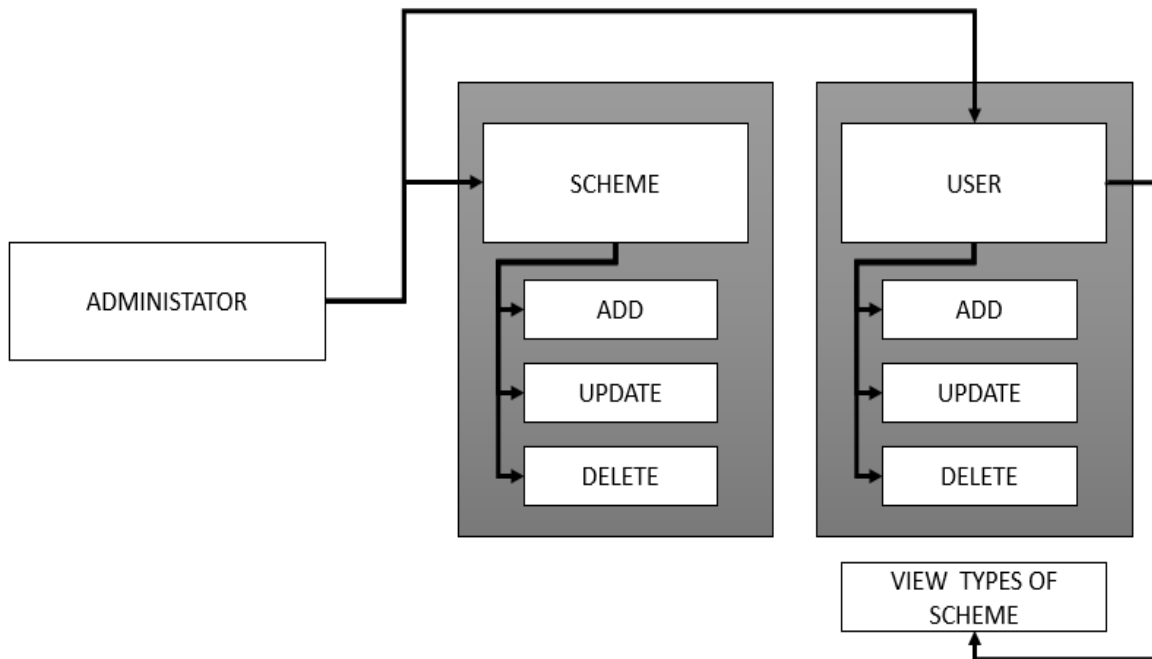


Fig.7 System Architecture

CHAPTER 8

SYSTEM TESTING

8.1 TEST PLAN

Software testing is the process of evaluation a software item to detect differences between given input and expected output. Also to assess the feature of a software item. Testing assesses the quality of the product. Software testing is a process that should be done during the development process. In other words software testing is a verification and validation process.

8.2 VERIFICATION

Verification is the process to make sure the product satisfies the conditions imposed at the start of the development phase. In other words, to make sure the product behaves the way we want it to.

8.3 VALIDATION

Validation is the process to make sure the product satisfies the specified requirements at the end of the development phase. In other words, to make sure the product is built as per customer requirements.

8.4 BASICS OF SOFTWARE TESTING

There are two basics of software testing: black box testing and white box testing.

8.5 BLACK BOX TESTING

Black box testing is a testing technique that ignores the internal mechanism of the system and focuses on the output generated against any input and execution of the system. It is also called functional testing.

8.6 WHITE BOX TESTING

White box testing is a testing technique that takes into account the internal mechanism of a system. It is also called structural testing and glass box testing. Black box testing is often used for validation and white box testing is often used for verification.

8.7 TYPES OF TESTING

There are many types of testing like

- 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 the testing of an individual unit or group of related units. It falls under the class of white box testing. It is often done by the programmer to test that the unit he/she has implemented is producing expected output against given input.

➤ **Integration Testing**

Integration testing is testing in which a group of components are combined to produce output. Also, the interaction between software and hardware is tested in integration testing if software and hardware components have any relation. It may fall under both white box testing and black box testing.

➤ **Functional Testing**

Functional testing is the testing to ensure that the specified functionality required in the system requirements works. It falls under the class of black box testing.

➤ **System Testing**

System testing is the testing to ensure that by putting the software in different environments (e.g., Operating Systems) it still works. System testing is done with full system implementation and environment. It falls under the class of black box testing.

➤ **Stress Testing**

Stress testing is the testing to evaluate how system behaves under unfavorable conditions. Testing is conducted at beyond limits of the specifications. It falls under the class of black box testing.

➤ **Performance Testing**

Performance testing is the testing to assess the speed and effectiveness of the system and to make sure it is generating results within a specified time as in performance requirements. It falls under the class of black box testing.

➤ **Usability Testing**

Usability testing is performed to the perspective of the client, to evaluate how the GUI is user-friendly? How easily can the client learn? After learning how to use, how proficiently can the client perform? How pleasing is it to use its design? This falls under the class of black box testing.

➤ **Acceptance Testing**

Acceptance testing is often done by the customer to ensure that the delivered product meets the requirements and works as the customer expected. It falls under the class of black box testing.

➤ Regression Testing

Regression testing is the testing after modification of a system, component, or a group of related units to ensure that the modification is working correctly and is not damaging or imposing other modules to produce unexpected results. It falls under the class of black box testing

8.8 REQUIREMENT ANALYSIS

Requirement analysis, also called requirement engineering, is the process of determining user expectations for a new modified product. It encompasses the tasks that determine the need for analyzing, documenting, validating and managing software or system requirements. The requirements should be documentable, actionable, measurable, testable and traceable related to identified business needs or opportunities and define to a level of detail, sufficient for system design.

8.8.1 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 often 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.

8.8.2 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 9

CONCLUSION

By using ML and DM techniques the prediction of scholarship much easier and effective also. While predicting the data we need to look upon many parameters. ML and DM techniques unable to work without training data. For collecting of datasets makes more difficult. We cannot be able to collect the best algorithm suitable for prediction. Because prediction is based on training data values: By using the ML and DM methods, the instructions can easily track on the scholarships and the problems faced by the recipients in their studies.

REFERENCES

- [1] Thalia Anagnos, Eva Schiorring, "Helping Students Get More than Their Money from an Engineering Scholarship Program" IEEE, 2018.
- [2] Angela R. Bielefeldt, "Characteristics of Engineering Faculty Engaged in the Scholarship of Teaching and Learning" IEEE, 2015.
- [3] K. Bunkar, "Data Mining: Prediction for Performance Improvement of Graduate Students using," IEEE, 2012.
- [4] Sercan, Saatci, Hande cansiz, gulsha aslan,erkan ozhan "Artificial intelligence-based Scholarship DQG Credit pre-assessment system" IEEE, 2017.
- [5] G. Upton, "The Effects of Merit-Based Scholarships on Educational Outcomes," Journal of Labor Research, vol. 37, no. 2, pp. 235-261.
- [6] Okfalisa, Ratik a fitriani, Yelli Vitriana, " The Comparison of Linear Regression Method and k- Nearest Neighbors in Scholarship Recipient" IEEE, 2018

APPENDICES

SOURCE CODE

```
package in.government.plans.identifier;

import java.math.BigDecimal;

import java.time.LocalDate;

import org.springframework.boot.CommandLineRunner;

import org.springframework.boot.SpringApplication;

import org.springframework.boot.autoconfigure.SpringBootApplication;

import org.springframework.context.annotation.Bean;

import org.springframework.scheduling.annotation.EnableScheduling;

import in.government.plans.identifier.models.AmmaVodiEligibility;

import in.government.plans.identifier.models.Description;

import in.government.plans.identifier.models.FeeReimbursementEligibility;

import in.government.plans.identifier.models.Registration;

import in.government.plans.identifier.models.Scheme;

import in.government.plans.identifier.models.YsrRaithuBharosaEligibility;

import in.government.plans.identifier.repository.AmmaVodiRepository;

import in.government.plans.identifier.repository.DescriptionRepository;

import in.government.plans.identifier.repository.FeeReimbursementRepository;

import in.government.plans.identifier.repository.RegistrationRepository;

import in.government.plans.identifier.repository.SchemeRepository;
```

```

import in.government.plans.identifier.repository.YsrRaithuBharosaRepository;

@SpringBootApplication

@EnableScheduling

public class GovernmentPlansIdentifierApplication {

    public static void main(String[] args) {

        SpringApplication.run(GovernmentPlansIdentifierApplication.class,
args);

    }

    @Bean

    public CommandLineRunner loadRegistrationData(RegistrationRepository
repository) {

        return (args) -> {

            repository.save(new Registration(null, "admin FN", "admin LN",
"admin@admin.com", LocalDate.now(), "admin",

                "admin"));

            repository.save(

                new Registration(null, "user FN", "user LN",
"user@user.com", LocalDate.now(), "user", "user"));

        };

    }

    @Bean

    public CommandLineRunner loadDescriptionData(DescriptionRepository
repository) {

```

```

return (args) -> {
    repository.save(new Description(null, "Amma vodi",
        "Jaganna Amma Vodi Scheme is one of the welfare
schemes among "
        + "the nine schemes of YSRCP
government called as “Navratanal”. "
        + "Amma Vodi Scheme was the part
of the election manifesto of "
        + "the present YSRCP government of
the state. It was introduced "
        + "by the Chief Minister YSR Jagan
Mohan Reddy on 10th June 2019. "
        + "The implementation of the scheme
is scheduled to be done on " + "9th January 2020."));

```

```

    repository.save(new Description(null, "YSR Rythu Bharosa",
        "The Government of Andhra Pradesh is
implementing, “YSR RYTHU "
        + "BHAROSA” from 15th October
2019 to provide financial "
        + "assistance to the farmer families
including tenant farmers "
        + "across the state @ Rs.12,500/- per
farmer family per year to "
        + "support the cultivators in meeting
the investment during the "

```

```

+ "crop season with a view to enable
them to "

+ "timely sourcing of quality inputs
and services for higher " + "crop productivity."));

repository.save(new Description(null, "Fee Reimbursement",

"Fee Reimbursement Scheme (also called as Post-
matric "

+ "Scholarship Scheme) is a student
education sponsorship Programme "

+ "by Government of Andhra Pradesh.
It supports students belonging to "

+ "economically weaker sections in
the state. In 2012-13, more than 600,000 students "

+ "in professional colleges were
covered under the scheme, including around "

+ "150,000 students in engineering
colleges. From 2013 to 2014, "

+ "the government has made Aadhaar
number mandatory for students availing the "

+ "scholarship."));

};

}

@Bean

```

```

    public CommandLineRunner loadSchemaData(SchemeRepository repository)
    {

        return (args) -> {

            repository.save(new Scheme(null, "YSR Rythu Bharosa",
            "Andhra Pradesh", "India", "Y.S. Jagan Mohan Reddy"));

            repository.save(new Scheme(null, "Fee Reimbursement",
            "Andhra Pradesh", "India", "Y.S. Jagan Mohan Reddy"));

            repository.save(new Scheme(null, "Amma vodi", "Andhra
            Pradesh", "India", "Y.S. Jagan Mohan Reddy"));

        };

    }

    @Bean

    public CommandLineRunner
    loadYsrRaithuBharosaData(YsrRaithuBharosaRepository repository) {

        return (args) -> {

            repository.save(new YsrRaithuBharosaEligibility(null, "Andhra
            Pradesh", "Agriculture", 5));

            repository.save(new YsrRaithuBharosaEligibility(null, "Andhra
            Pradesh", "Farmer", 5));

        };

    }

    @Bean

```

```

public                               CommandLineRunner
loadFeeReimbursementData(FeeReimbursementRepository repository) {

    return (args) -> {

        repository.save(new           FeeReimbursementEligibility(null,
BigDecimal.valueOf(120000), false, "SC",
                                   BigDecimal.valueOf(200000), false));

        repository.save(new           FeeReimbursementEligibility(null,
BigDecimal.valueOf(120000), false, "ST",
                                   BigDecimal.valueOf(200000), false));

        repository.save(new           FeeReimbursementEligibility(null,
BigDecimal.valueOf(120000), false, "ST",
                                   BigDecimal.valueOf(200000), true));

        repository.save(new           FeeReimbursementEligibility(null,
BigDecimal.valueOf(120000), false, "ST",
                                   BigDecimal.valueOf(200000), true));

        repository.save(new           FeeReimbursementEligibility(null,
BigDecimal.valueOf(120000), false, "EBC",
                                   BigDecimal.valueOf(100000), true));

        repository.save(new           FeeReimbursementEligibility(null,
BigDecimal.valueOf(120000), false, "BC",
                                   BigDecimal.valueOf(150000), false));

        repository.save(new           FeeReimbursementEligibility(null,
BigDecimal.valueOf(120000), false, "BC",
                                   BigDecimal.valueOf(150000), true));
    }
}

```

```

        };

    }

    @Bean

    public CommandLineRunner loadAmmaVoditData(AmmaVodiRepository
repository) {

        return (args) -> {

            repository.save(new AmmaVodiEligibility(null, "School", true,
true));

            repository.save(new AmmaVodiEligibility(null, "College", true,
true));

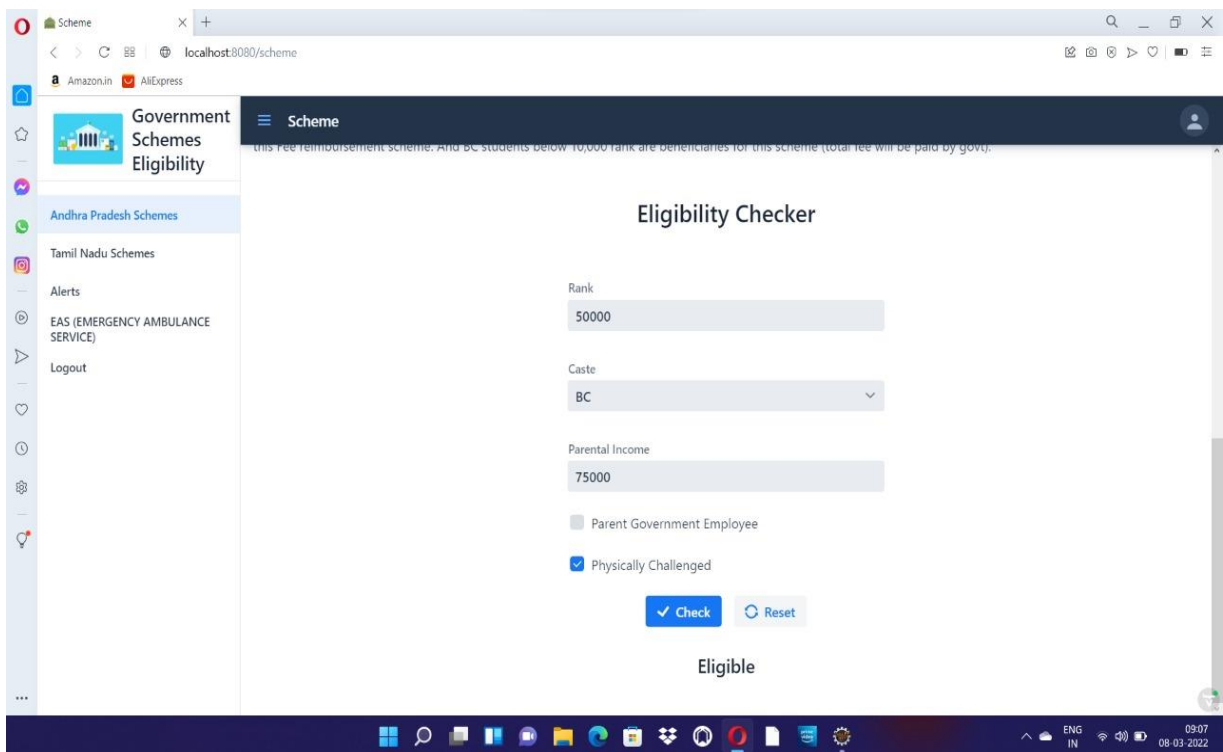
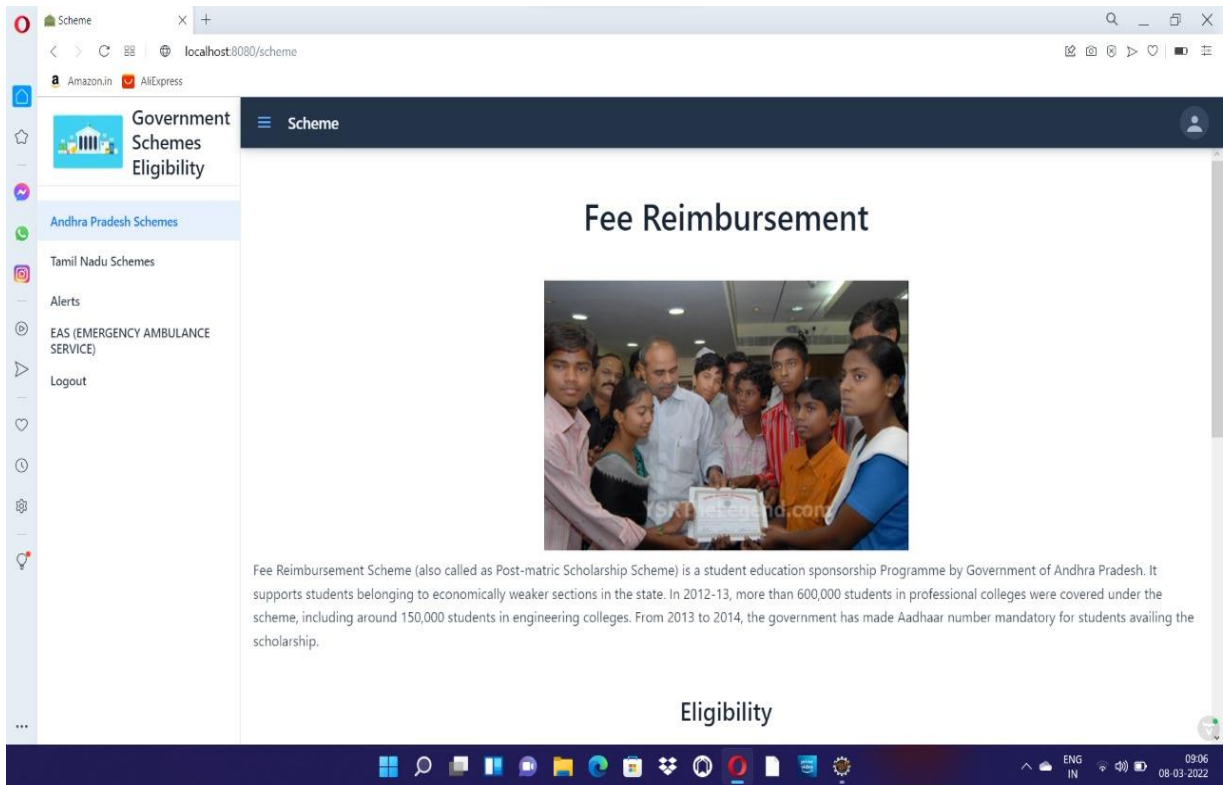
        };

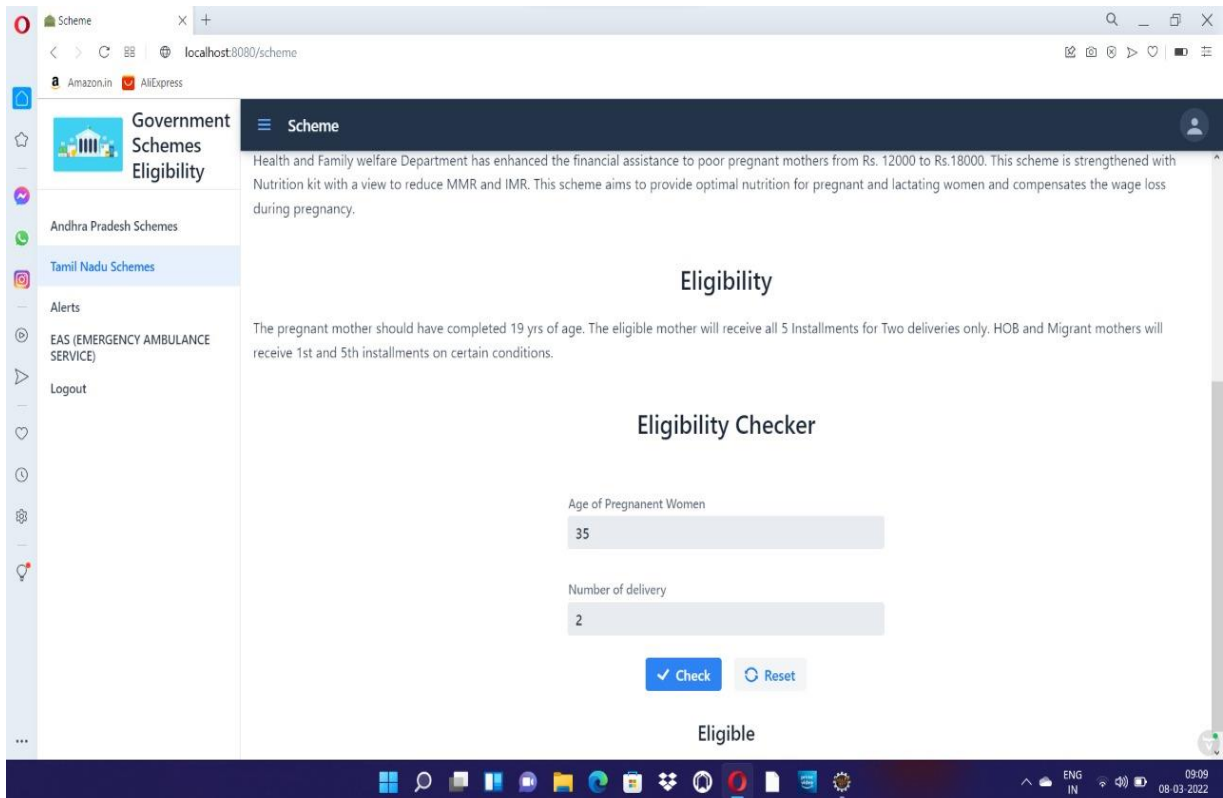
    }

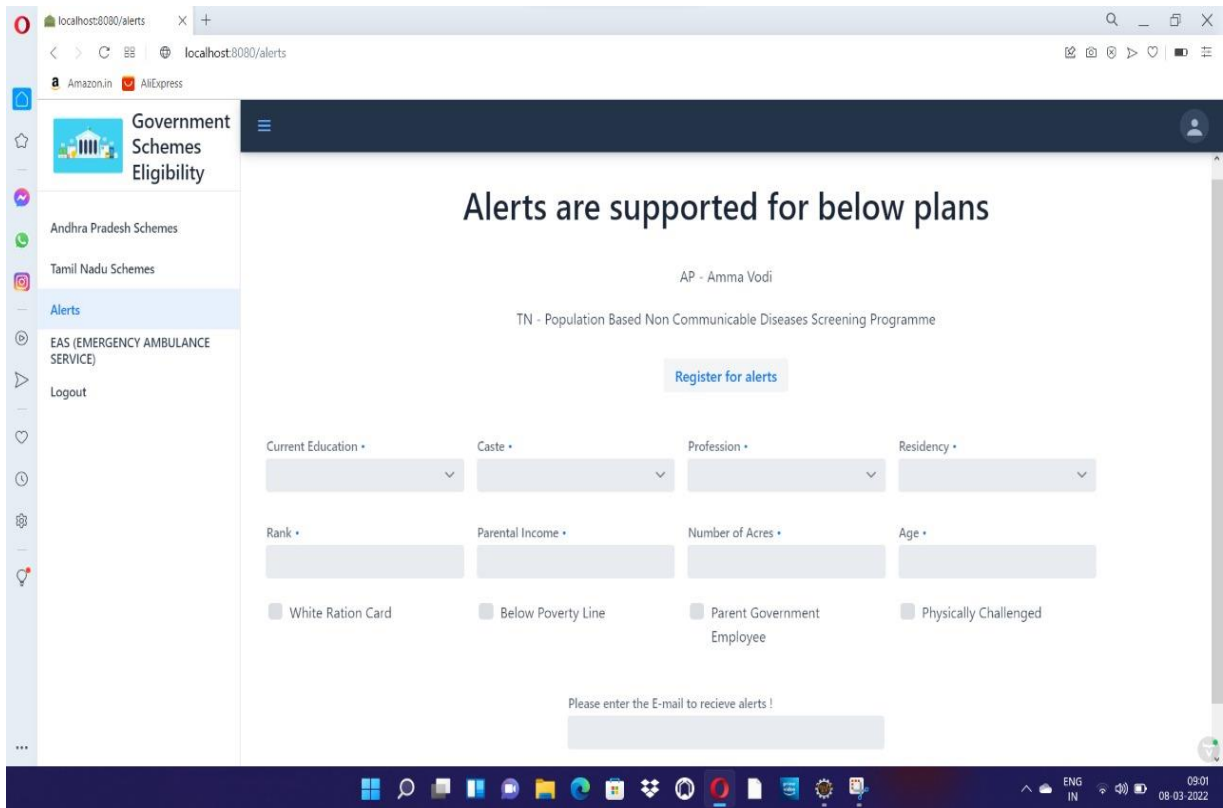
}

```

SCREEN SHOTS







A REVIEW ON DATA MINING AND MACHINE LEARNING METHODS FOR STUDENT SCHOLARSHIP PREDICTION

Yadla Venkatesh
Dept of CSE
Sathyabama Institute of
Science and Technology,
Chennai, India.
venkateshyadla282@gmail.com

Pachigolla Rakesh
Dept of CSE
Sathyabama Institute of
Science and Technology,
Chennai, India.
pachigollarakesh@gmail.com

Dr. M.Saravanan M.E., Ph.D.,
Assistant Professor,
Dept of CSE,
Sathyabama Institute of
Science and Technology,
Chennai, India.
saravanan.cse@sathyabama.ac.in

ABSTRACT

Numerous administration subsidized projects are decreased to clean on the grounds that the specialists don't track down the perfect individuals. In this way, there is a dependable and precise method for deciding if a hesitant individual has the option to apply for a program, to screen the preparation interaction, to have a youngster, or to have a solicitation or refusal from the specialists. Our framework will remember more subtleties for the administrator framework. He can likewise look further into how to utilize it. The individual might acknowledge or dismiss the client care plan. The business can likewise see the idea of the application interaction at their level.

Keywords- Machine Learning, Scholarship Prediction

INTRODUCTION

School achievement influences understudy advancing as well as monetarily. In Indonesia, the failure to pay educational expenses is one of the significant difficulties for school graduates, particularly in non-administrative colleges. Hence, bursaries from government, establishments, establishments, and so forth are expected to help these understudies. Garut High School of Technology, Garut School of Technology (later STTG) is one of the private establishments where 1,018 understudies concentrate most during the 2016/2017 school year. A portion of these understudies have financial and social issues, however their high grades require a four year college education. In such manner, consistently STTG gives subsidizing to different associations, including public and private ventures and tasks. Be that as it may, because of the set number of grants, there is a contrast between the quantity of understudies

qualified for the grant and the quantity of understudies required. To defeat this, STTG settles on a solid decision to move the grant to one side. In such manner, the execution of ICT will assist with making genuine, quick and proficient choices. An enormous number of understudies are offered grants, which is a component of understudy data.

OBJECTIVE

The reason for the framework is to foster a framework that remembers government projects and data for recipients. The objective is to set up a framework that starts important cycles and permits clients to play out an assortment of exercises. The objective is to set up a framework that starts important cycles and permits clients to play out an assortment of exercises.

SCOPE OF THE PROJECT

The size of the state program is as per the following: Notification of the state program is not difficult to enroll and utilize, and to save time. The size of an administration call is to check out

straightforwardness and security. The best-in-class program assists clients with utilizing the framework and screen their status consistently. Through this site, the client assists with deciding whether the person has the privilege to the natural product, or then again in the event that the specialists will not acknowledge their solicitation. This site assists everybody with controlling the level they apply to their program.

LITERATURE SURVEY

This cycle is tedious and furthermore monotonous. Then again, it is now and then truly challenging to settle on genuine choices. The utilization of specialized abilities and AI abilities to decide grant beneficiaries is significant for deliberate and functional appraisals of candidates. The data utilized in this study was gathered on the web, and the motivation behind this study was to address an inquiry created by Kredi ve Yurtlar Kurumu and to decide the specialized abilities of the grant college understudies. also, machine preparing is deliberate and intentional in a suitable manner [4]. The review utilized information extraction methods, associating that the creators with the yearly meeting report of the American Society for Engineering Education (ASEE) would recognize development educators working at SoTL. The ASEE record writers' arrangement was distributed in 2012-2014 on Learning Through Service (LTS), specialized morals, and plan, and is the writer of an article on development, assembling, the climate, and power. and PC parts The ASEE report found the segment qualities of development educators across the country. [2] Scholarships can be a significant wellspring of help and consolation for college understudies, yet they are not an assurance of ideal graduation or graduation. Understudies will be impacted by your own, wellbeing, and monetary

issues, and they will stagger in school and not be able to apply for grants. To lessen the effect of the grant, the NSF program, supported by the S-STEM Engineering Leadership Pathway Scholars (ELPS), gives financing and exercises to empower and get ready first year recruits for their four-year college education. certificate in designing, morals, information and abilities to keep on being a forerunner in the 21st century [1]. This prescience will assist you with tracking down feeble understudies and get passing marks. C4.5, ID3, and CART tree choice calculations are utilized to decide the quantity of designing understudies to decide the exhibition of a last test of the year. The result of the choice tree anticipated the quantity of understudies who might win, lose, or advance one year from now. The outcomes give steps to work on the presentation of understudies who have fizzled or are relied upon to progress. After the consequences of the last, most important test were declared, the understudies entered the food they were eating into the framework and dissected it in ensuing classes. A similar examination of the outcomes proposes that the expectations assisted understudies with handicaps gain ground and further develop their results [3]. Highlights utilized incorporate semester, GPA, work letter, assist letter, understudy ID, ID with carding, family card, preparing card, note, financial balance, and letter of credit. The factors utilized in the correlation technique incorporate True, Note, Memory, Sorting Error, Complete Error, and Root Mean Square Error. The data given by 8212 grant beneficiaries is checked against 90:10, 70:30, 50:50, 30:70, and 10:90. Here, Rapidminer is utilized as a device to see scientific outcomes from the two sides. Accordingly, both the 90:10 and 70:30 data modes give 100 percent exactness, accuracy, and memory [6].

EXISTING SYSTEM

Today, different government programs are distributed in the media and distributed by important associations. It is hard for individuals to get data about help accessibility. The objectives of the program are regularly not met because of helpless correspondence between the different divisions and directors. Individuals don't have the foggiest idea about their arrangements and don't profit from them.

PROPOSED SYSTEM

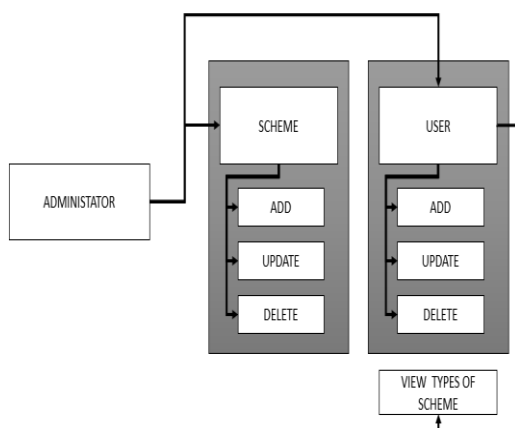
The framework is a product-based site that stores a public data set of all program-related data. It's not difficult to track down the data you need. Managers and the overall population can sign in to the framework and apply online for planned projects. This framework permits the coordination of all degrees of associations and related associations. The framework additionally takes into consideration ordinary observing and data trade.

ADVANTAGES OF PROPOSED SYSTEM

It gives every one of the projects reported by the public authority to general society.

- another arrangement is reported.
- Display promotions on the principle page.
- The enrollment cycles.
- Login.
- Allows clients to control their data.

ARCHITECTURE DIAGRAM



PROPOSED ALGORITHM

K-NEAREST NEIGHBOR (KNN) CLASSIFICATION METHOD

K-NN is a model in view of preparing that carries execution to the grassroots level and postpones counting until it is executed. The k-NN calculation is the easiest in AI calculations. Neighbors from a line of things whose classes (on k-NN classifications) or worth of property (on k-NN backwardness) are known.

Stage 1: Start

Stage 2: Input: $D = \{(x_1, c_1), \dots, (x_N, c_N)\}$

Stage 3: another instance of arranging $x = (x_1 \dots x_N)$

Stage 4: Count (x_i, c_i) $d(x_i, x)$ for each case wherein it is composed.

Stage 5: Separate $d(x_i, x)$ from base to top, $(I = 1 \dots N)$

Stage 6: x : Select K for instance close to $D_k x$

Stage 7: $D_k x$ is doled out to classification x , which has on numerous occasions

Stage 8: Conclusion

MODULES

Administrator

User

Schemes

Report

Result

IMPLEMENTATION

ADMINISTRATOR

Head The executive will need to check for refreshes, then, at that point, update all the

data on the administrator site and send a SMS to the chairman assuming the client agrees with the program. The Director will check the qualification data and the arrangement will show this data to individuals whenever they have applied. The administrator is answerable for exercises connected with the public authority gave program. This module centers around getting various applications for clients.

USER

Assuming the client can't peruse the program, the individual first visits the site, then, at that point, really looks at the prerequisites for the program, and afterward gets a message that the person in question can apply for the program. They additionally have a simple method for getting to the framework and screen the application structure. Just client data can be refreshed or erased from the rundown.

SCHEMES

This module permits you to utilize administrator, and the administrator can add new projects, erase past projects, and update past projects. They show every one of the subtleties of the program, just as the shrewd client, the models and prerequisites. Just the administrator can make another plan on the web. Any remaining clients can get to the subtleties of the program

REPORT

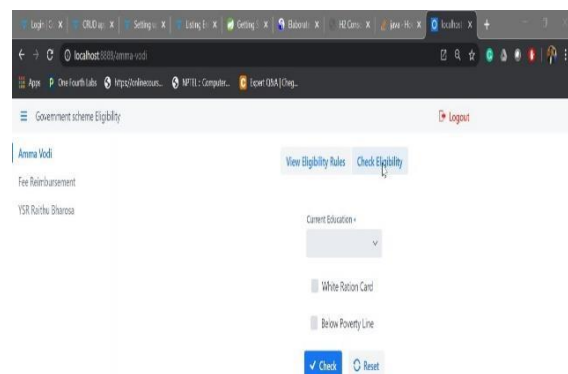
In the report module, we can check out different reports, for example, the knowledge report framework and the insight report clients. Different reports for clients, for example, legitimate prerequisites, application pausing, and dismissal. The blueprint of the report incorporates the name of the picture, the kind of plot, the personality card, and the end date. The client explicit report should incorporate all client data, itemized data about the necessities and prerequisites.

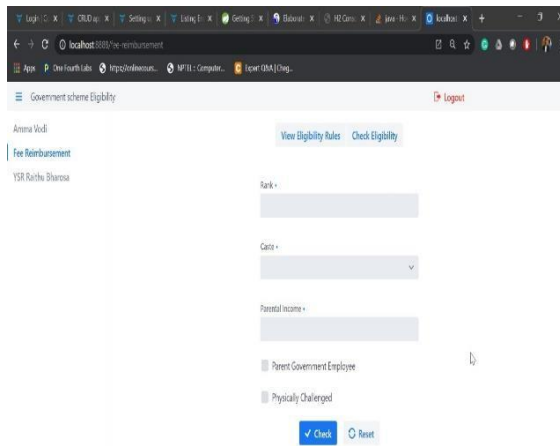
Subtleties of the program and every one of the thoughts will be introduced in the authority report. All dismissed application reports will be remembered for the application report

RESULT

To offer this support to applicable clients, we initially request that clients actually look at all the data on every one of the above things, and afterward subsequent to gathering client data utilizing the enlistment structure for their clients, we will actually want to offer an extraordinary assistance to the individual or clients concerned. By utilizing it, customers can profit from government programs.

SCREENSHOTS





CONCLUSION

By using ML and DM techniques the prediction of scholarship much easier and effective also. While predicting the data we need to look upon many parameters. ML and DM techniques unable to work without training data. For collecting of datasets makes more difficult. We cannot be able to collect the best algorithm suitable for prediction. Because prediction is based on training data values: By using the ML and DM methods, the instructions can easily track on the scholarships and the problems faced by the recipients in their studies.

REFERENCES

- [1] Thalia Anagnos, Eva Schiorring, "Helping Students Get More than Their Money from an Engineering Scholarship Program" IEEE, 2018.
- [2] Angela R. Bielefeldt, "Characteristics of Engineering Faculty Engaged in the Scholarship of Teaching and Learning" IEEE, 2015.
- [3] K. Bunkar, "Data Mining: Prediction for Performance Improvement of Graduate Students using," IEEE, 2012.
- [4] Sercan, Saatci, Hande cansiz, gulsha aslan, erkan ozhan "Artificial intelligence-based Scholarship DQG Credit pre-assessment system" IEEE, 2017.
- [5] G. Upton, "The Effects of Merit-Based Scholarships on Educational Outcomes," Journal of Labor Research, vol. 37, no. 2, pp. 235-261.
- [6] Okfalisa, Ratik a fitriani, Yelli Vitriana, "The Comparison of Linear Regression Method and k- Nearest Neighbors in Scholarship Recipient" IEEE, 2018
- [7]. Roger Pressnma, "SOFTWARE ENGINEERING", 7 th edition
- [8]. Yehuda Shiran: Java Script Programming, 2008.
- [9]. Holzner: HTML Black Book (HTML4)
- [10]. Patel moss: Java Database Programming with JDBC
- [11]. J2EE Professional by Shadab siddiqui
- [12]. Nick Todd: JAVA server pages
- [13]. Larne Pekowsley: JAVA server pages

A REVIEW ON DATA MINING AND MACHINE LEARNING METHODS FOR STUDENT SCHOLARSHIP PREDICTION (1).docx

ORIGINALITY REPORT

1 %

SIMILARITY INDEX

1 %

INTERNET SOURCES

1 %

PUBLICATIONS

0 %

STUDENT PAPERS

PRIMARY SOURCES

1

www.researchgate.net

Internet Source

1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On