

USAGE OF MACHINE LEARNING FOR STRATEGIC DECISION MAKING AT HIGHER EDUCATIONAL INSTITUTIONS

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

by

Yasam Kavya (38110666)
Gudimetla Anjali(38110175)



**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 with "A" grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600119

www.sathyabama.ac.in



DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **YasamKavya(38110666)** and **Gudimetla Anjali(38110175)** who carried out the project entitled "**usage of machine learning for strategic decision making higher educational institutions**" under my supervision from November 2021 to March 2022.

Internal Guide

Dr. B. Bharathi M.E., Ph.D.

Head of the Department

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I **YasamKavya (Reg No:38110666)** and **Gudimetla Anjali(Reg No: 38110175)** hereby declare that the Project Report entitled **“usage of machine learning for strategic decision making higher educational institutions”** done by us under the guidance of **Dr. B. BharathiM.E., Ph.D.** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in 2018-2022.

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.S.Vigneshwari M.E., Ph.D.** and **Dr.L.Lakshmanan M.E., Ph.D.** , Heads of the Department of Computer Science and Engineering for providing us necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and a deep sense of gratitude to my Project Guide **Dr. B. Bharathi M.E., Ph.D.**, for her 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

In many organizations, machine learning techniques are used for analyzing large amount of available data's, information's for decision making process. In educational sector, Machine learning is used for wide variety of applications such as suggestion to the students based on 10th mark and interest. One of the most important milestones in an individual's life involves self-analysis, critical thinking and finally decision making. In this paper, we present survey results on which academic decisions they concern and the variables involved in them. Using machine learning algorithms, we predicted courses/institutions in a real case study to support decision making. The choice of the career is influenced by views of your parents, friends, relatives, teachers and the media. Today with a wider choice and an ever increasing competition, you need to plan your career wisely and at the earliest. While choosing a stream after 10th, a training course or a career and 12th groups you should know your abilities, interests, and personality. Besides these you should gather information regarding different career options, the eligibility criteria, the premier institutions/Schools, and other criteria of selection and the market demands. The present system undertaken by Education Department, on different educational and vocational courses available and the institutions through competitive examination after 10th and 12th aims at giving you the much needed information.

TABLE OF CONTENTS

Chapter No.	TITLE	Page No.
	ABSTRACT	v
	LIST OF FIGURES	viii
	LIST OF ABBREVIATIONS	ix
1	INTRODUCTION	11
	1.1. OBJECTIVE	12
	1.2. PROPOSED ALGORITHM	12
	1.2.1.KNN ALGORITHM	12
	1.2.2.K-MEANS CLUSTERING ALGORITHM	17
2	LITERATURE SURVEY	19
	2.1. RELATED WORK	19
3	METHODOLOGY	23
	3.1. SYSTEM REQUIREMENTS	
	3.1.1.HARDWARE REQUIRMENTS	23
	8	

3.1.2.SOFTWARE REQUIREMENTS	23
3.2.SOFTWARE ENVIRONMENT	23
3.2.1.APPLICATION OF JAVA	23
3.2.2.MYSQL	27
3.3.FEASIBILITY STUDY	28
3.4. EXISTING SYSTEM	29
3.5.PROPOSED SYSTEM	30
3.6.SYSTEM DESIGN	30
3.6.1.INPUT DESIGN	30
3.6.2.OUTPUT DESIGN	31
3.5.	31
3.7.UML DIAGRAMS	31
3.7.1.USE CASE DIAGRAM	32
3.7.2.SEQUENCE DIAGRAM	33
3.7.3.COLLABORATION DIAGRAM	34
3.8.MODULES	34
3.8.1.DATA COLLECTION MODULE	35
3.8.2.PRE-PROCESSING MODULE	36
3.8.3.FEATURE EXTRACTION	36
3.8.4.APPLY MACHINE LEARNINGALGORITHE	36

	3.8.5.PREDICTION MODULE	37
	3.9.SYSTEM ARCHITECTURE	37
4	RESULTS AND DISCUSSION	39
	4.1.FUNCTIONAL REQUIREMENTS	39
	4.2.NON-FUNCTIONAL REQUIREMENTS	40
		41
5	CONCLUSION	42
		42
	REFERENCE	42
	APPENDICES	
	A.SOURCE CODE	44
	B. SCREENSHOTS	44

LIST OF FIGURES

Figure No.	Figure Name	Page No.
1.2.1.	KNN CLASSIFIER	3
1.2.1.	KNN GRAPH	4
1.2.2.	K-MEANS CLUSTERING	8
3.2.1.	PLATFORM INDEPENDENT DIAGRAM	15
3.2.1.	COLLECTION FRAME WORK	17

3.7.1.	USE CASE DIAGRAM	23
3.7.2.	SEQUENCE DIAGRAM	24
3.7.3.	COLLABORATION DIAGRAM	27
3.9	SYSTEM ARCHITECTURE	28

LIST OF ABBREVIATIONS

ABBREVIATIONS	EXPANSION
ML	Machine Learning
KNN	K nearest neighbour
MYSQL	My sequel

RDBMS

Relational database management
system

OS

Operating system

CHAPTER 1

INTRODUCTION

Two imperatives for better use of data confront higher education. The first is driven by external factors while the second is driven internally by continuous quality improvement. Steep declines in financial and public support have driven efforts by governments to collect data that support the proposition that institutions are accountable for the revenue they receive. Working from a defensive posture, many colleges and universities have been able to waylay undesirable changes by satisfying external requests for data. At a higher level, however, those institutions that deliberately use data to improve overall performance meet compliance-based requirements while enacting a future that is informed by data. The proposition that higher education's approach to data use has changed very little may be disputed. At the same time, it also is clear that technology has made new conversations possible. New techniques including analytics or predictive analytics provide institutions new opportunities to use data to improve their efficiency while better serving students (see, for example, Bichsel, 2012 and WCET, n.d.). Colleges and universities are entering an era in which strategic information about student learning and success, budgeting, and efficiency can be united under the umbrella of big data. Higher education is now collecting more data than ever before. However, these efforts are most often directed at the first imperative, compliance reporting, rather than the second imperative, improving institutional strategy. Forward thinking institutions will quickly resolve this seeming dichotomy. They will seek opportunities to build capacity, remove constraints to span existing boundaries that determine data use and find ways to bring data and strategy together. The result can advance institutional mission, meeting external policy demands and improving student success. Strategic thinking and the data that serve those strategies come at a price. In this chapter, we review both opportunities and barriers associated with creating and using actionable strategic and operational data. We also identify successful steps for data use based on our experiences in working with higher education institutions to facilitate strategic planning and to create cultures of inquiry and evidence. We also survey emerging technologies and their promise to help institutions help their students. This chapter is intended to provide

practical advice and not to provide a theoretical overview of the tenets of strategic planning. Institutions sufficiently courageous to engage in a data journey require support. Toward that end, this chapter also provides advice drawn from personal experience and new developments in management science to help navigate these new pathways.

1.1. OBJECTIVE

One of the main objects of the recently developed policy for higher education regards the creation of a more diversified higher education system with flexible and adaptive institutions. In educational sector, machine learning is used for wide variety of applications such as suggestion to the students based on 10th mark and interest. Hoping that it would help you in understanding what each course for career entails and help you introspect and analyze the required abilities to be successful, happy and content in the course/career you choose. While choosing a stream after 10th, a training course or a career and 12th groups you should know your abilities, interests, and personality. Besides these you should gather information regarding different career options, the eligibility criteria, the premier institutions/Schools, and other criteria of selection and the market demands.

1.2. PROPOSED ALGORITHMS

1.2.1. KNN ALGORITHMS

- o K-Nearest neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- o K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

- o K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- o K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- o K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.
- o It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- o KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.
- o **Example:** Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.



Fig 1.2.1: KNN classifier

Why do we need a K-NN Algorithm?

Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x_1 , so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. Consider the below diagram:

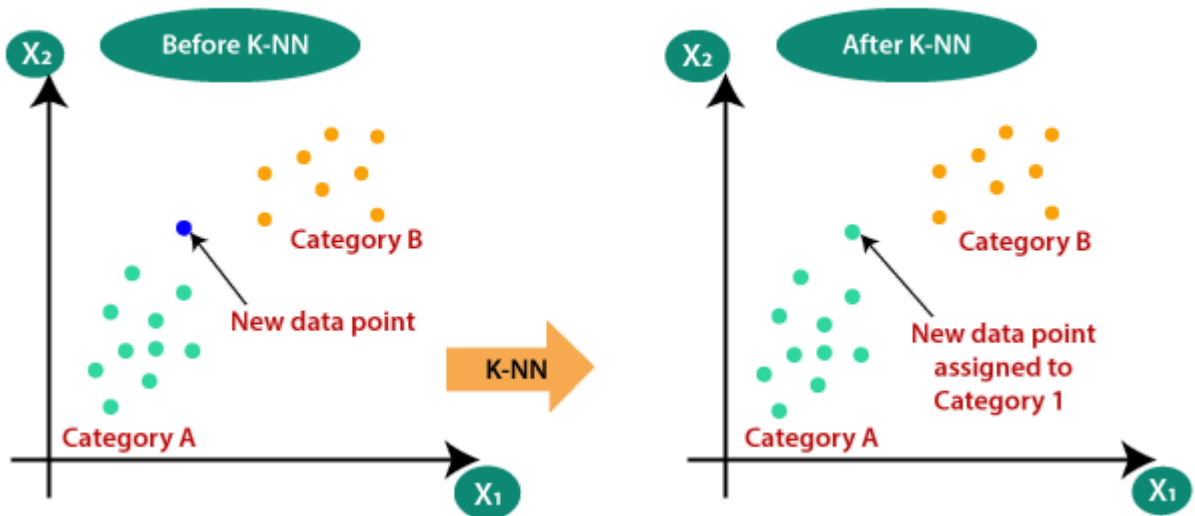


Fig 1.2.1 KNN graph

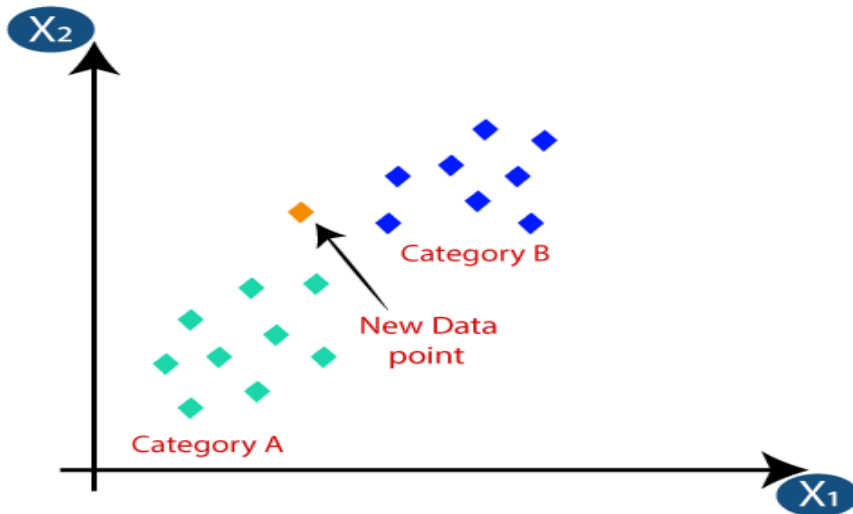
How does K-NN work?

The K-NN working can be explained on the basis of the below algorithm:

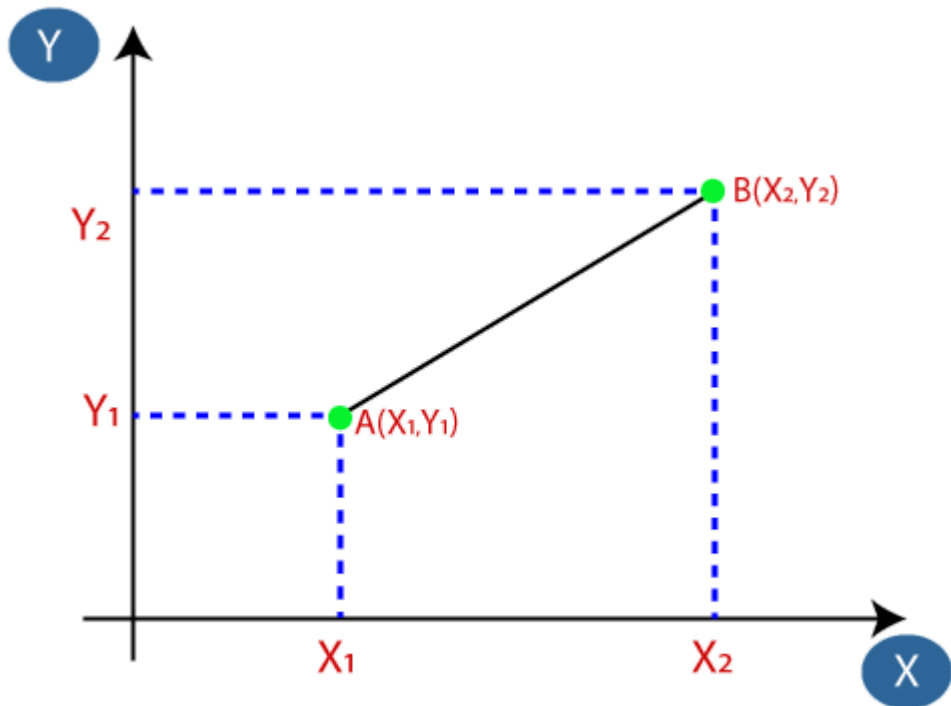
- o **Step-1:** Select the number K of the neighbors
- o **Step-2:** Calculate the Euclidean distance of **K number of neighbors**
- o **Step-3:** Take the K nearest neighbors as per the calculated Euclidean distance.
- o **Step-4:** Among these k neighbors, count the number of the data points in each category.

- o **Step-5:** Assign the new data points to that category for which the number of the neighbor is maximum.
- o **Step-6:** Our model is ready.

Suppose we have a new data point and we need to put it in the required category.
Consider the below image:



- o Firstly, we will choose the number of neighbors, so we will choose the $k=5$.
- o Next, we will calculate the **Euclidean distance** between the data points. The Euclidean distance is the distance between two points, which we have already studied in geometry. It can be calculated as:



Euclidean Distance between A_1 and $B_2 = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$

- o By calculating the Euclidean distance we got the nearest neighbors, as three nearest neighbors in category A and two nearest neighbors in category B. Consider the below image:



- o As we can see the 3 nearest neighbors are from category A, hence this new data point must belong to category A.

How to select the value of K in the K-NN Algorithm?

Below are some points to remember while selecting the value of K in the K-NN algorithm:

- o There is no particular way to determine the best value for "K", so we need to try some values to find the best out of them. The most preferred value for K is 5.
- o A very low value for K such as $K=1$ or $K=2$, can be noisy and lead to the effects of outliers in the model.
- o Large values for K are good, but it may find some difficulties.

Advantages of KNN Algorithm:

- o It is simple to implement.

- o It is robust to the noisy training data
- o It can be more effective if the training data is large.

Disadvantages of KNN Algorithm:

- o Always needs to determine the value of K which may be complex some time.
- o The computation cost is high because of calculating the distance between the data points for all the training samples.

K-NEAREST NEIGHBOR(KNN) CLASSIFICATION METHOD

K-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is among the simplest of all machine learning algorithms. The neighbors are taken from a set of objects for which the class (for k-NN classification) or the object property value (for k-NN regression) is known.

STEP 1: BEGIN

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

STEP 3: $x = (x_1, \dots, x_n)$ new instance to be classified

STEP 4: FOR each labelled instance (x_i, c_i) calculate $d(x_i, x)$

STEP 5: Order $d(x_i, x)$ from lowest to highest, $(i = 1 \dots N)$

STEP 6: Select the K nearest instances to x: D_x^k

STEP 7: Assign to x the most frequent class in D_x^k

STEP 8: END

1.2.2. K-MEANS CLUSTERING ALGORITHM:

What is K-means?

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

Input: k (the number of clusters),
 D (a set of lift ratios)
Output: a set of k clusters
Method:
Arbitrarily choose k objects from D as the initial cluster centers;
Repeat:
 1. (re)assign each object to the cluster to which the object is
 the most similar, based on the mean value of the objects in the
 cluster;
 2. Update the cluster means, i.e., calculate the mean value of
 the objects for each cluster
Until no change;

Fig 1.2.2 : k means clustering

Details of K-means

1. Initial centroids are often chosen randomly. - Clusters produced vary from one run to another
2. The centroid is (typically) the mean of the points in the cluster.
3. 'Closeness' is measured by Euclidean distance, cosine similarity, correlation, etc.
4. K-means will converge for common similarity measures mentioned above.
5. Most of the convergence happens in the first few iterations. - Often the stopping condition is changed to 'Until relatively few points change clusters'

CHAPTER 2

LITERATURE SURVEY

2.1 RELATED WORK

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.

Barriers and drivers of innovation in higher education: Case study-based evidence across ten European Universities

The paper advances current knowledge on factors affecting higher education institutions in their quest for innovation in education. Based on an analysis of ten institutional cases from five European countries, a comprehensive description and classification of barriers and drivers of innovation are provided. Results indicate certain “disengagement” in relation of higher education institutions and education policy makers, business, and students as well as between higher education institutions’ managers and their subordinates. Based on the findings, major innovation-related challenges in the higher education are discussed and related practical recommendations are presented.

Supporting academic decision making at higher educational institutions using machine learning-based algorithms

Decisions made by deans and university managers greatly impact the entire academic community as well as society as a whole. In this paper, we present survey results on which academic decisions they concern and the variables involved in them. Using machine learning algorithms, we predicted graduation rates in a real case study to support decision making. Real data from five undergraduate engineering programs at District University Francisco Jose de Caldas in Colombia illustrate our results. The comparison between support vector machine and artificial neural network is held using the confusion matrix and the receiver operating characteristic curve. The algorithm methods and architecture are presented.

The effect of blended learning on student performance at course-level in higher education: A meta-analysis

The present paper analyzes the impact of blended learning (BL) on the academic achievement of higher education students. A meta-analysis ($k = 51$ effect sizes) was conducted to perform a statistical synthesis of studies contrasting student performance in BL conditions with traditional classroom instruction. We include disciplines and instructors' end-of-course evaluation method as moderating variables. The results show that BL demonstrates a small summary effect ($g^+ = 0.385$, $p < 0.001$) compared to traditional teaching methods. A significantly higher mean effect size was found in STEM disciplines ($g^+ = 0.496$) compared to that of non-STEM disciplines ($g^+ = 0.210$). Nevertheless, the weighted mean effect sizes reveal no significant differences regarding end-of-course assessment methods, namely one-moment and multiple-component assessment. The finding confirms that BL is significantly associated with greater learning performance of STEM-disciplined students than with traditional classroom practice. Accordingly, discussion concerning the findings and implications for future research are elaborated.

Using system dynamics to develop education for sustainable development in higher education with the emphasis on the sustainability competencies of students

In response to growing concerns of the community about sustainability challenges and the intensification of the international calls to move towards a sustainable future, higher education should be involved in implementing the programs of education for sustainable development; because of this fact that University graduates are part of this solution as future leaders and inheritors of technology. In this study, the underlying research question is: What mechanisms are needed to develop education for sustainable development in higher education with the emphasis on the sustainability competencies of students? The research method was system dynamics. Therefore, we used a mix method research design. Tools of research included the literature review, questionnaire, interview, and observation. We developed a dynamic model to develop the education for sustainable development in higher education with the emphasis on the sustainability competencies of students. This model describes

the research problem and predicts the behavior of model variables by simulating in the next 20 years. This model included 18 reinforcing and six balancing feedback loops. After ensuring the validity of the model, mechanisms were elicited from the model. Finally, we evaluated these mechanisms for finding the impacts on improving the problem.

Strategic alignment between sustainability and information systems: A case analysis in Malaysian public higher education Institutions

Higher Education Institutions have embedded sustainability initiatives into their core activities of curriculum, research, community, and operational to respond the global transformation towards a sustainable future. Numerous studies have been conducted concerning an integrated sustainability into the missions and strategies of the Higher Education Institutions. However, there is a lack of works that highlight the important role of information systems to support the sustainability practices in Higher Education Institutions. This problem motivates the study to underscore the important role of information systems during the sustainability implementation. A preliminary case analysis is performed to observe how the Higher Education Institutions implement the sustainability initiatives and to proof the knowledge gap in real practice. The analysis was conducted during visits via a semi-structured interview with a number of persons from Higher Education Institutions in Malaysia that have been implementing sustainable campus. According to the preliminary case analysis, the Higher Education Institutions do not contemplate the information system strategy at the beginning of their sustainability implementation. There is a lack of information system application to support their sustainability practices. They still use manual methods to collect the sustainability metrics and assess their sustainability performance. As a result, the sustainability decision-making is isolated and they cannot evaluate their sustainable performance. The problems of managing the sustainability data and processes arise. Therefore, a concept of the strategic alignment between sustainability and information system is underscored in the present study. The introduction of this concept through this study would enhance our understanding in term of the significant role of the

information systems to support the sustainability practices and to achieve a successfully sustainable campus.

CHAPTER 3

METHODOLOGY

3.1. SYSTEM REQUIREMENTS

3.1.1. HARDWARE REQUIREMENTS:

- System - Pentium-IV
- Speed - 2.4GHZ
- Hard disk - 40GB

- Monitor - 15VGA color
- RAM - 512MB

3.1.2. SOFTWARE REQUIREMENTS:

- Operating System - Windows XP
- Coding language - Java
- IDE - Net beans
- Database -MySQL

3.2. SOFTWARE ENVIRONMENT

INTRODUCTION

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 waned with passage of time, while Java's has grown.

3.2.1. APPLICATION OF JAVA

Java is widely used in every corner of world and of human life. Java is not only used in softwares 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:

1. Developing Desktop Applications
2. Web Applications like LinkedIn.com, Snapdeal.com etc
3. Mobile Operating System like Android
4. Embedded Systems

5. Robotics and games etc.

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.

Eg: Pointers and Operator Overloading are not there in java but were an important part of C++.

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.

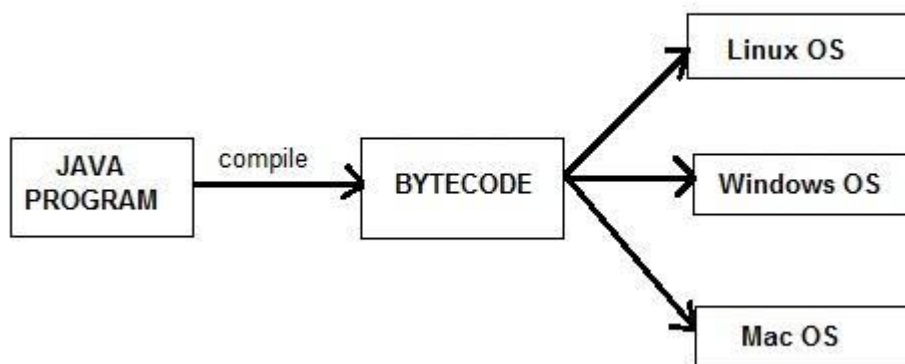


Fig 3.2.1: platform independent diagram

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

9) 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.

COLLECTION 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.

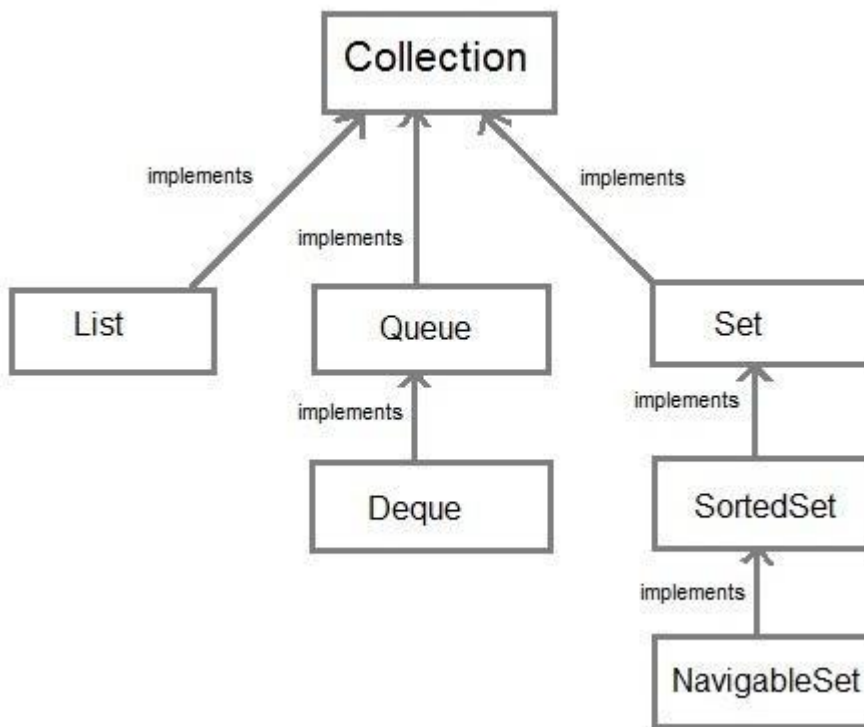


Fig 3.2.1 : collection framework

3.2.2. MYSQL

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, Perl/PHP/Python." Freesoftware-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.

3.3. 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.

ECONOMICAL FEASIBILITY

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

TECHNICAL FEASIBILITY

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

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity.

3.4. EXISTING SYSTEM

In existing system, three supervised classification algorithms are deployed to predict graduation rates from real data about undergraduate engineering students in South America. The analysis of receiver operating characteristic curve and accuracy are executed as measures of effectiveness to compare and evaluate decision tree, logistic regression, and random forest, where this last one demonstrates the best outcomes.

DISADVANTAGE OF THE EXISTING SYSTEM

- Existing system miss the undesirable data for the students.
- And It may not check the social data for the student.

3.5. PROPOSED SYSTEM

In proposed system, machine learning is used for wide variety of applications such as suggestion to the students based on 10th mark and interest. One of the most important milestones in an individual's life involves self-analysis, critical thinking and finally decision making. While choosing a stream after 10th, a training course or a career and 12th groups you should know your abilities, interests, and personality. Besides these you should gather information regarding different career options, the eligibility criteria, the premier institutions/Schools, and other criteria of selection and the market demands.

ADVANTAGES OF PROPOSED SYSTEM

- Accuracy is high.
- High efficiency

3.6. SYSTEM DESIGN

3.6.1. 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.

3.6.2. 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 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

3.7. 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.

3.7.1. 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.

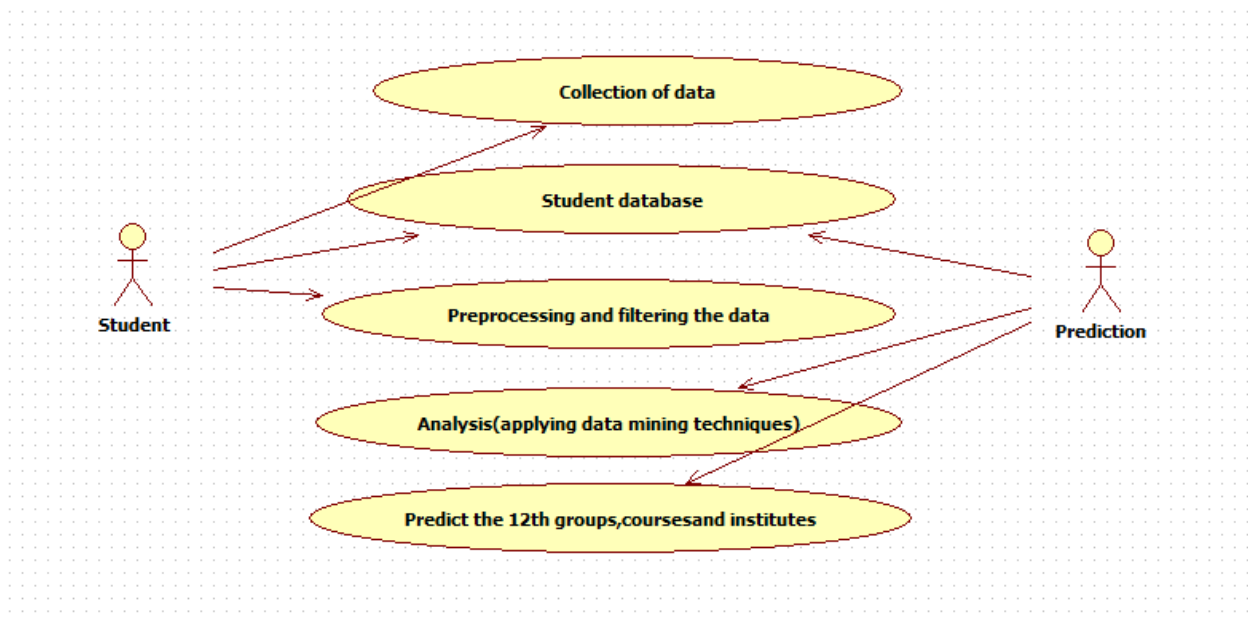


Fig 3.7.1 : use case diagram

3.7.2. SEQUENCE DIAGRAM

Sequence diagram model the flow of logic within your system in a visual manner, enabling you both to document and validate your logic, and commonly used for both analysis and design purpose. Sequence diagram are the most popular UML artifact for dynamic modeling, which focuses on identifying the behavior within your system.

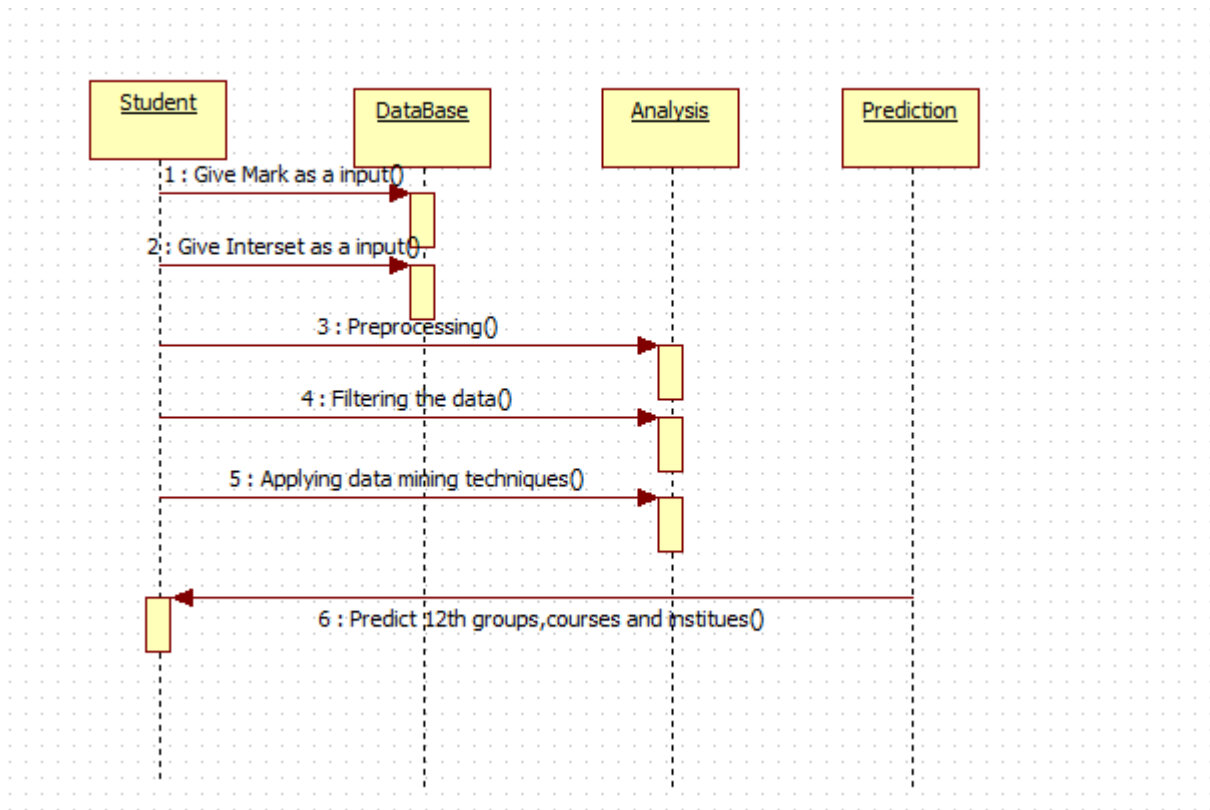


Fig 3.7.2 : sequence diagram

3.7.3. COLLABORATION DIAGRAM

Another type of interaction diagram is the collaboration diagram. A collaboration diagram represents a collaboration, which is a set of objects related in a particular context, and interaction, which is a set of messages exchange among the objects within the collaboration to achieve a desired outcome.

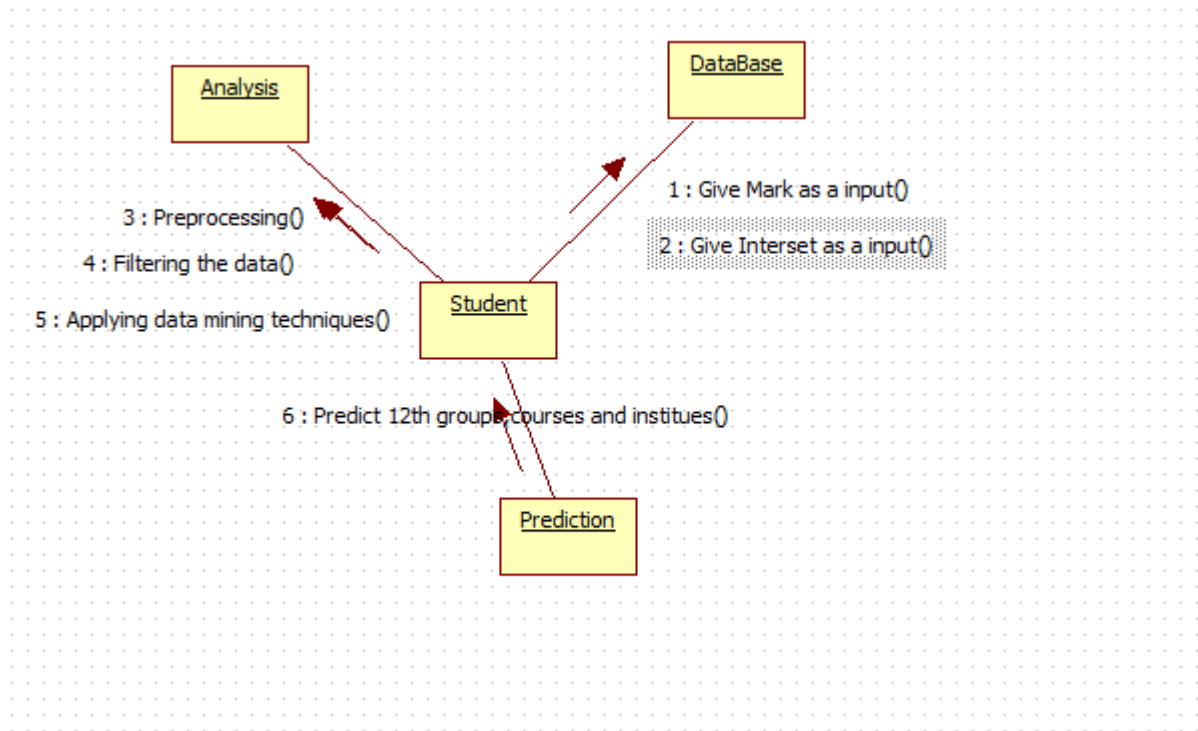


Fig 3.7.3 : collaboration diagram

3.8. MODULES

- Data Collection Module
- Pre-Processing Module
- Feature Extraction Module
- Apply Machine Learning Algorithm

- Prediction

3.8.1. DATA COLLECTION MODULE:

Data is a collection global dataset. IN this system use Pima Indian data set is used for training a model. The training data is a initial set of data which is used to understand the program. This is the one in which we have to train the model first because to set the feature and this data is available on system. This data is used to teach the machine for do different actions. It is the data in which model can learn with algorithm to teach the model and doing work automatic. Testing data is the input given to a software. It shows the data affects when the execution of the module that specifying and this is basically used for testing.

3.8.2. PRE-PROCESSING MODULE:

- Data preprocessing is a process in which that is actual use for converting the basic data into the clean data set. It is the step in which the data transform or an encode to the state that the machine can be easily parse. The major task of data preprocessing in learning process is to remove the unwanted data and filling the missed value. So that it help to machine can be trained easily.

3.8.3. FEATURE EXTRACTION

- Feature Extraction is the method in which it used for alter the key data for features of outcomes. This, trait square is used to compute the characteristics of designs given that facilitate in different amid the class of key pattern details. This method involving to decrease the counts of resource required to describe the huge set of data. Feature
- Extraction is an attribute reduction process. This is also used to increasing the speed and effectiveness of supervised learning.

3.8.4. APPLY MACHINE LEARNING ALGORITHMS

- The ML algorithm is the non-parametric method proposed by Thomas Cover used for Regression and Classification. This algorithm is mainly used for the classification of problems in the industry. Machine Learning algorithm is a type

of instance-based learning method. This algorithm relies on the distance for objects classification, training data normalizing to the improve its accuracy dramatically. The neighbors are derived from the set of things for which classes or object property values are known. It can be thought of as a training set for the algorithm, although no explicit training steps are required.

3.8.5. PREDICTION MODULE:

- In this module, machine learning is used for wide variety of applications such as suggestion to the students based on 10th mark and interest. One of the most important milestones in an individual's life involves self-analysis, critical thinking and finally decision making. While choosing a stream after 10th, a training course or a career and 12th groups you should know your abilities, interests, and personality. Besides these you should gather information regarding different career options, the eligibility criteria, the premier institutions/Schools, and other criteria of selection and the market demands.

3.9. 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.

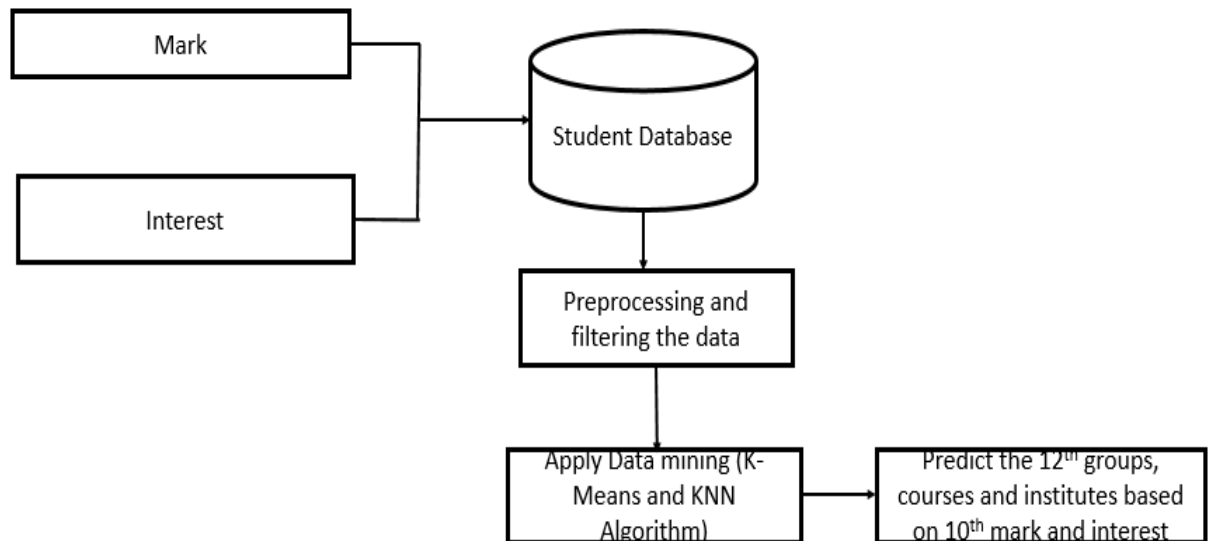


Fig 3.9 : System architecture

CHAPTER 4

RESULTS AND DISCUSSION

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.

4.1. 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.

4.2. 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 5

CONCLUSION

In this paper, the classification task is used on student database to predict the student's decision on the basis of previous database. As there are many approaches that are used for data classification, the k-means clustering Classifier and KNN Classifier are used here. Information's like marks, interest were collected from the student's previous database, to predict the training course or a career and 12th groups at the end. This study will help to the students to improve the decision of the student. This can help the students improve in their academics, which eventually leads to a good decision in their career.

REFERENCES

- [1] A. Lašáková, L. Bajžíková, and I. Dedze, "Barriers and drivers of innovation in higher education: Case study-based evidence across ten European Universities," *Int. J. Educ. Develop.*, vol. 55, pp. 69–79, May 2017.
- [2] Y. V. N. Acevedo and C. E. M. Marín, "System architecture based on learning analytics to educational decision makers toolkit," *Adv. Comput. Sci. Eng.*, vol. 13, no. 2, pp. 89–105, 2014.
- [3] A. R. T. Góes, M. T. A. Steiner, and P. J. S. Neto, "Education quality measured by the classification of school performance using quality labels," *Appl. Mech. Mater.*, vols. 670–671, pp. 1675–1683, Oct. 2014.
- [4] H. Lounis and T. Fares, "Using efficient machine-learning models to assess two important quality factors: Maintainability and reusability," in *Proc. Joint Conf. 21st Int. Workshop Softw. Meas. 6th Int. Conf. Softw. Process Product Meas.*, Nov. 2011, pp. 170–177.

- [5] Y. Nieto, V. García-Díaz, C. Montenegro, and R. G. Crespo, "Supporting academic decision making at higher educational institutions using machine learning-based algorithms," *Soft Comput.*, vol. 23, no. 12, pp. 4145–4153, 2018.
- [6] K. V. Pincus, D. E. Stout, J. E. Sorensen, K. D. Stocks, and R. A. Lawson, "Forces for change in higher education and implications for the accounting academy," *J. Accounting Educ.*, vol. 40, pp. 1–18, Sep. 2016.
- [7] A.-P. Pavel, A. Fruth, and M.-N. Neacsu, "ICT and e-learning—Catalysts for innovation and quality in higher education," *Procedia Econ. Finance*, vol. 23, pp. 704–711, Jan. 2015.
- [8] F. V. Elena, A. M. Manuel, and G. G. S. Carina, "Which use give teachers at La Laguna University to ICTs?" *Procedia-Social Behav. Sci.*, vol. 93, pp. 1646–1651, Oct. 2013.
- [9] H. M. Vo, C. Zhu, and N. A. Diep, "The effect of blended learning on student performance at course-level in higher education: A meta-analysis," *Stud. Educ. Eval.*, vol. 53, pp. 17–28, Jun. 2017.
- [10] I. M. Taucean and M. Tamasila, "Research challenges for eLearning support in engineering and management training," *Procedia-Social Behav. Sci.*, vol. 124, pp. 210–218, Mar. 2014.
- [11] E. Faham, A. Rezvanfar, S. H. M. Mohammadi, and M. R. Nohooji, "Using system dynamics to develop education for sustainable development in higher education with the emphasis on the sustainability competencies of students," *Technol. Forecast. Social Change*, vol. 123, pp. 307–326, Oct. 2017.
- [12] F. A. Goni, A. G. Chofreh, M. Mukhtar, S. Sahran, S. A. Shukor, and J. J. Klemeš, "Strategic alignment between sustainability and information systems: A case analysis in Malaysian public higher education Institutions," *J. Clean. Prod.*, vol. 168, pp. 263–270, Dec. 2017.
- [13] I. González-González and A. I. Jiménez-Zarco, "Using learning methodologies and resources in the development of critical thinking competency: An

exploratory study in a virtual learning environment,” Comput. Hum. Behav., vol. 51, pp. 1359–1366, Oct. 2015.

APPENDICES

A) SOURCE CODE

Login view

```
.login-view {  
  
    display: grid;  
    padding: 30%  
}  
  
:host {
```

```
        display: block;

        padding: var(--lumo-space-m) var(--lumo-space-l);

    }
}
```

Prediction client

```
package com.example.application.client;

import org.springframework.cloud.openfeign.FeignClient;
import org.springframework.web.bind.annotation.PostMapping;
import org.springframework.web.bind.annotation.RequestBody;

import com.example.application.data.entity.BtechSubjects;
import com.example.application.data.entity.IntermediateSubjects;
import com.example.application.data.entity.SchoolSubjects;

@FeignClient(name = "prediction-client", url = "localhost:1080")
public interface PredictionClient {

    @PostMapping("predict")
    public PredictionResponse predict(@RequestBody SchoolSubjects features);

    @PostMapping("predict1")
    public PredictionResponse predict1(@RequestBody IntermediateSubjects features);

    @PostMapping("predict2")
    public PredictionResponse predict2(@RequestBody BtechSubjects features);
}
```

```
}
```

Prediction response

```
packagecom.example.application.client;
```

```
public class PredictionResponse {
```

```
    private String prediction;
```

```
    public String getPrediction() {
```

```
        return prediction;
```

```
    }
```

```
    public void setPrediction(String prediction) {
```

```
        this.prediction = prediction;
```

```
    }
```

```
    @Override
```

```
    public String toString() {
```

```
        return "PredictionResponse [prediction=" + prediction + "];
```

```
    }
```

```
}
```

SCHOOL SUBJECTS

```
packagecom.example.application.data.entity;
```

```
import javax.persistence.Entity;
```

```
import com.example.application.data.AbstractEntity;
```

```
import java.time.LocalDate;
```

```
import javax.annotation.Generated;
```

```
@Entity
```

```
public class SchoolSubjects extends AbstractEntity {
```

```
    private Integer maths;
```

```
    private Integer biology;
```

```
    private Integer physics;
```

```
    private Integer chemistry;
```

```
    private Integer civics;
```

```
    private Integer economics;
```

```
    private Integer history;
```

```
    private SchoolSubjects(Builder builder) {
```

```
        this.maths = builder.maths;
```

```
        this.biology = builder.biology;
```

```
        this.physics = builder.physics;
```

```
        this.chemistry = builder.chemistry;
```

```
        this.civics = builder.civics;
```

```
        this.economics = builder.economics;

        this.history = builder.history;
    }
}
```

```
public SchoolSubjects() {
    super();
}
```

```
public Integer getMaths() {
    return maths;
}
```

```
public void setMaths(Integer maths) {
    this.maths = maths;
}
```

```
public Integer getBiology() {
    return biology;
}
```

```
public void setBiology(Integer biology) {
    this.biology = biology;
}
```

```
public Integer getPhysics() {  
    return physics;  
}
```

```
public void setPhysics(Integer physics) {  
    this.physics = physics;  
}
```

```
public Integer getChemistry() {  
    return chemistry;  
}
```

```
public void setChemistry(Integer chemistry) {  
    this.chemistry = chemistry;  
}
```

```
public Integer getCivics() {  
    return civics;  
}
```

```
public void setCivics(Integer civics) {  
    this.civics = civics;  
}
```

```

public Integer getEconomics() {
    return economics;
}

```

```

public void setEconomics(Integer economics) {
    this.economics = economics;
}

```

```

public Integer getHistory() {
    return history;
}

```

```

public void setHistory(Integer history) {
    this.history = history;
}

```

@Override

```

public String toString() {
    return "SchoolSubjects [maths=" + maths + ", biology=" + biology + ",
    physics=" + physics + ", chemistry="
        + chemistry + ", civics=" + civics + ", economics=" + economics +
    ", history=" + history + "];"
}

```

/**

```
* Creates builder to build {@link SchoolSubjects}.
```

```
* @return created builder
```

```
*/
```

```
public static Builder builder() {  
    return new Builder();  
}
```

```
/**
```

```
* Builder to build {@link SchoolSubjects}.
```

```
*/
```

```
public static final class Builder {  
    private Integer maths;  
    private Integer biology;  
    private Integer physics;  
    private Integer chemistry;  
    private Integer civics;  
    private Integer economics;  
    private Integer history;  
  
    private Builder() {  
    }  
  
    public Builder withMaths(Integer maths) {  
        this.maths = maths;  
    }  
}
```

```
        return this;
    }
}
```

```
public Builder withBiology(Integer biology) {
    this.biology = biology;
    return this;
}
```

```
public Builder withPhysics(Integer physics) {
    this.physics = physics;
    return this;
}
```

```
public Builder withChemistry(Integer chemistry) {
    this.chemistry = chemistry;
    return this;
}
```

```
public Builder withCivics(Integer civics) {
    this.civics = civics;
    return this;
}
```

```
public Builder withEconomics(Integer economics) {
```

```
        this.economics = economics;
        return this;
    }

    public Builder withHistory(Integer history) {
        this.history = history;
        return this;
    }

    public SchoolSubjects build() {
        return new SchoolSubjects(this);
    }
}
```

B) SCREENSHOTS

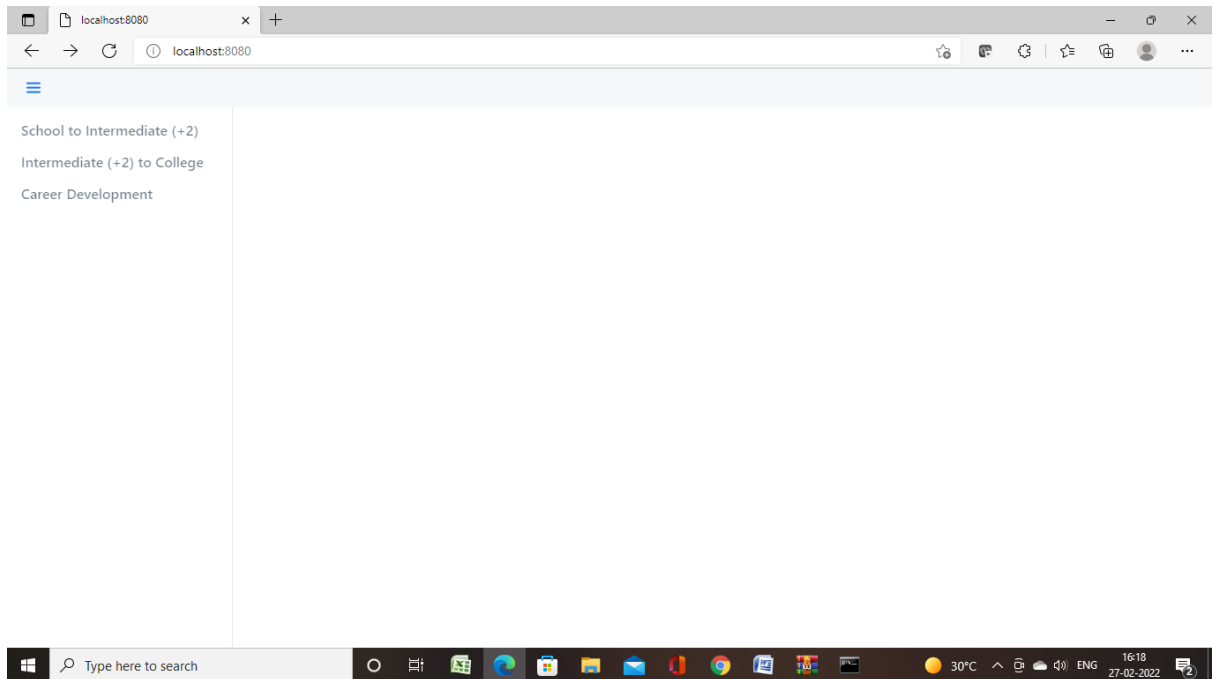


Fig: home page

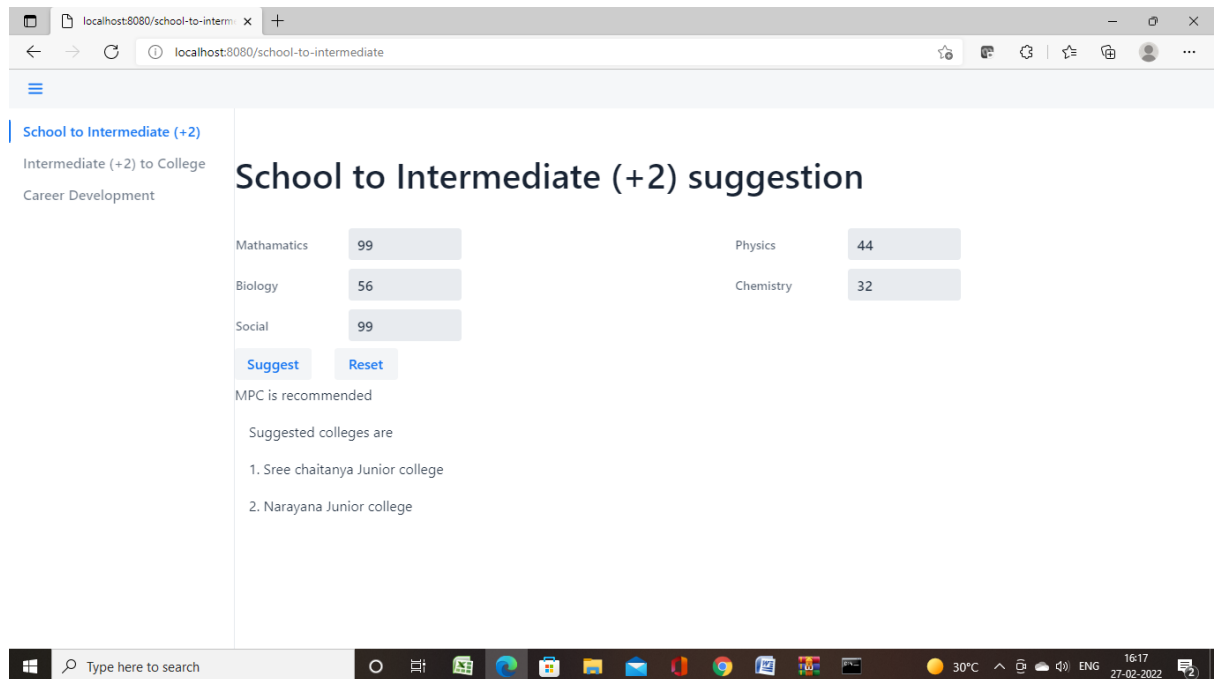


Fig: tenth-twelfth prediction

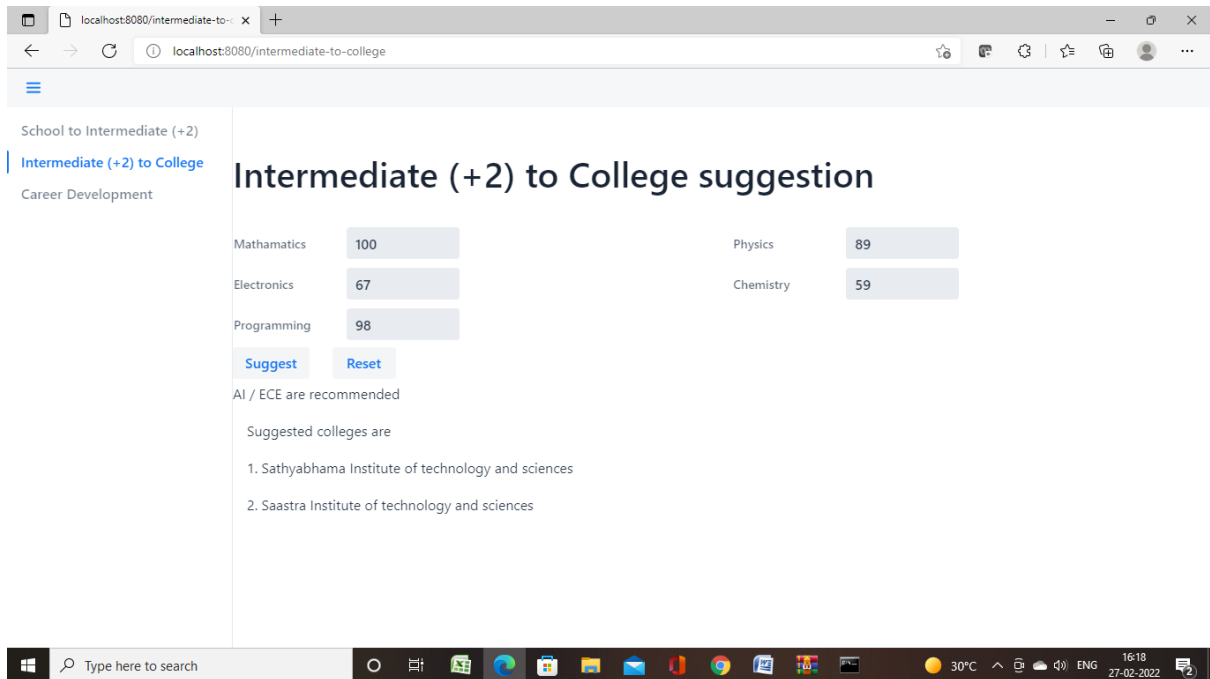


Fig: intermediate to college prediction

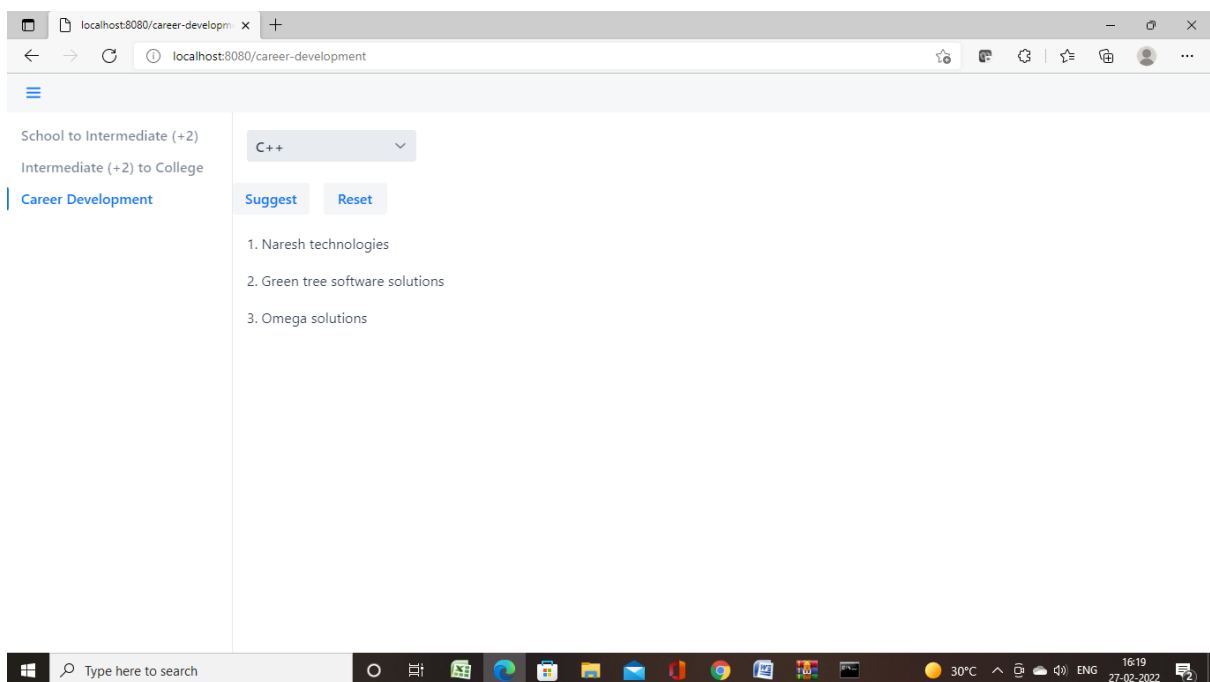


Fig: career prediction