HIGHER EDUCATION RECOMMENDATION SYSTEM USING MACHINE LEARNING ALGORITHMS

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

By

BARRI DINESH (Reg.No - 391100137)
KALAGA LOKANADHAM (Reg.No – 39110433)

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 | 12B Status by UGC | Approved by AICTE
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI - 600119

APRIL - 2023

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A" 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 **BARRI DINESH (Reg.No - 39110137) and KALAGA LOKANADHAM(Reg.No - 39110433)** who carried out the Project Phase-2 entitled **"HIGHER EDUCATION RECOMMENDATION SYSTEM USING MACHING ALGORITHMS"** under my supervision from January 2023 to April 2023.

Internal Guide

Dr. JEMSHIA MIRIAM M.E., Ph.D.

Head of the Department
Dr.L.LAKSHMANAN,M.E.,Ph.D.

Submitted for Viva voce Examination held on 24.04.2023

Internal Examiner

External Examiner

DECLARATION

I, **BARRI DINESH (Reg.No- 39110137),** hereby declare that the Project Phase-2 Report entitled **"HIGHER EDUCATION RECOMMENDATION SYSTEM USING MACHINE LEARNING"** done by me under the guidance of **Dr. JEMSHIA MIRIAM, M.E., Ph.D.** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in **Computer Science and Engineering**.

DATE:24.04.2023

PLACE: Chennai

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.,** Head 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. JEMSHIA MIRIAM M.E., Ph.D.,** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my phase-2 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 marks. One of the most important milestones in an individual life involves self-analysis, critical thinking and finally decision making. In this paper, we have presented a recommender system for undergraduate & graduate admission seekers, which can help students to choose best graduate university matching their academic profile. Here I have used a different data mining techniques to transform database of students of relevant information into a universal database format using academic data of successful students who have already got opportunity to study abroad. After that I have developed a machine learning algorithm which can calculate similarity between training and test data based on weighted scores. I have used K-nearest Neighbor algorithm and feature-weighted algorithm for calculating top N similar users for the test users and recommend Top K universities to users from N similar users. The choice of the career is influenced by views of your parents, friends, relatives, teachers, 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 an institution 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.

KEYWORDS: KNN Algorithm, Feature-weighted Algorithm...

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Many students who want to pursue higher studies apply different universities with their academic profile as well as standardized test scores such as SAT, GRE, TOEFL, and IELTS. Institutions take in the students who are suitable candidates based on their academic profile, standardized test scores. But in this entire process university selection is the most crucial & tedious step for applying to graduate studies. Some of them succeed and get admission into their desired programs in desired universities, but some are not because of the academic level of colleges which they have applied. To resolve this problem of not getting admission because of applications, even though students have good academic profile, I have developed this recommendation system. In this project, the knowledge acquired from the database of successful applicants is used to predict the schools with various data mining techniques. This data will be modeled into machine learning algorithms to predict the universities and their acceptance rate for the given user academic details. 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

For an aspiring student who wants to apply for higher studies in other countries, university selection process is a challenging task as lot of different criteria need to consider during application process based on individual's requirement. Some of them succeed and get admission into their desired programs in desired universities, but some are not because of the academic level of colleges, which they have applied. This problem can be addressed by modeling a recommender system based on various classification algorithms. In this project based on the student data set and the student profile who is looking for the admit, various models will be trained and a list of 10 best universities will be suggested such that it maximizes the chances of a student getting admit from that university list.

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.

2.1 INFERENCES FROM LITREATURE SURVEY

2.1.1 Author: Anna Lasakovaa, Lubica Bajzikovaa.

Title: Barriers and drivers of innovation in higher education: Case study-based evidence across ten European Universities- June 2017

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 recommended are presented.

2.1.2 Author: Hien M.Voab,Chang Zhaua

The effect of blended learning on student performance at course level in higher education: Ameta-analysis -June 2017

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 of 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 elaborate.

2.1.3 Author: Piotr Indyk,Yaron Singer,Ali Vakilian

Title: Supporting academic decision making at higher educational institutions using machine learning-based algorithms -August 2018

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.

2.1.4 Author: Monica Trakru, Tapan Kumar Jha

Title: Using system dynamics to develop education for sustainable development in higher education with the emphasis on the sustainability competencies of students -September 2019

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.

2.1.5 Author: Alcina Judy, Kesha D'cruz

Title: Strategic alignment between sustainability and information systems: A case analysis in Malaysian public higher education Institutions -July 2020

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.

2.2 OPEN PROBLEMS IN 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.
- It may not check the social data for the student.

CHAPTER 3

REQUIREMENT ANALYSIS

3.1 FEASIBILITY STUDIES/RISK ANALYSIS OF THE PROJECT

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

- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

3.1.1 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.

3.1.2 Social feasibility:

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

3.2 SOFTWARE REQUIREMENTS SPECIFICATION DOCUMENT

3.2.1 HARDWARE REQUIREMENTS:

 System - Pentium-IV
 Speed - 2.4GHZ
 Hard disk - 40GB
 RAM - 512MB

3.2.2 SOFTWARE REQUIREMENTS:

 Operating System – Windows 7 and greater versions
 Coding language – Python
 Frontend Technology- Flask
 IDE – Visual Studio Code

We can use any python compiler to run this project

Here is a step-by-step guide to installing Visual Studio Code and Python:

Installing Visual Studio Code:

- Go to the official Visual Studio Code website: https://code.visualstudio.com/
- Click on the "Download for Windows" button if you are using a Windows operating system, or click on the "Download for Mac" button if you are using a Mac.
- Once the download is complete, run the installation file and follow the prompts to install Visual Studio Code on your computer.

Installing Python:

- Go to the official Python website: https://www.python.org/
- Click on the "Downloads" link in the menu bar at the top of the page.
- Scroll down to the "Python Releases for Windows/MacOS/Linux" section and click on the version of Python you want to download (e.g. Python 3.10.0).

- Scroll down to the "Files" section and click on the appropriate installer for your operating system (e.g. "Windows x86-64 executable installer" for Windows 64-bit).
- Once the download is complete, run the installation file and follow the prompts to install Python on your computer.

Setting up Visual Studio Code for Python:

- Open Visual Studio Code.
- Click on the "Extensions" icon on the left-hand side of the window.
- Search for "Python" in the search bar.
- Click on the "Install" button next to the "Python" extension.
- Once the installation is complete, restart Visual Studio Code.
- Open a new file in Visual Studio Code and save it with a .py extension (e.g. "hello_world.py").
- Type in the following code:
- print("Hello, world!")
- Press Ctrl+Shift+P (Windows) or Command+Shift+P (Mac) to open the Command Palette.
- Type in "Python: Select Interpreter" and press Enter.
- Select the Python interpreter you installed earlier.
- Press F5 to run the code.
- You should see "Hello, world!" printed in the terminal window at the bottom of the Visual Studio Code window

.

Installing Modules

To install the modules you mentioned, open a command prompt or terminal window and enter the following commands:

- For flask, enter pip install flask
- For pandas, enter pip install pandas
- For numpy, enter pip install numpy
- For sklearn, enter pip install scikit-learn

CHAPTER 4

DESCRIPTION OF PROPOSED SYSTEM

In proposed system, machine learning is used for wide variety of applications such as suggestion to the students based on 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 institute or a training course or a 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.

4.1 SELECTED METHODOLOGY OR PROCESS MODEL

In our project we use waterfall model of our software development cycle because of its step-by-step procedure while implementing. Requirement Gathering and analysis – All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification document. System Design — The requirement specifications from first phase are studied in this phase and the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture. Implementation — With inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing. Integration and Testing – All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.

Deployment of system − Once the functional and non-functional testing is done; the product is deployed in the customer environment or released into the market. Maintenance − There are some issues which come up in the client environment. To fix those issues, patches are released. Also, to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer view.

4.2 ARCHITECTURE / OVERALL DESIGN OF PROPOSED SYSTEM

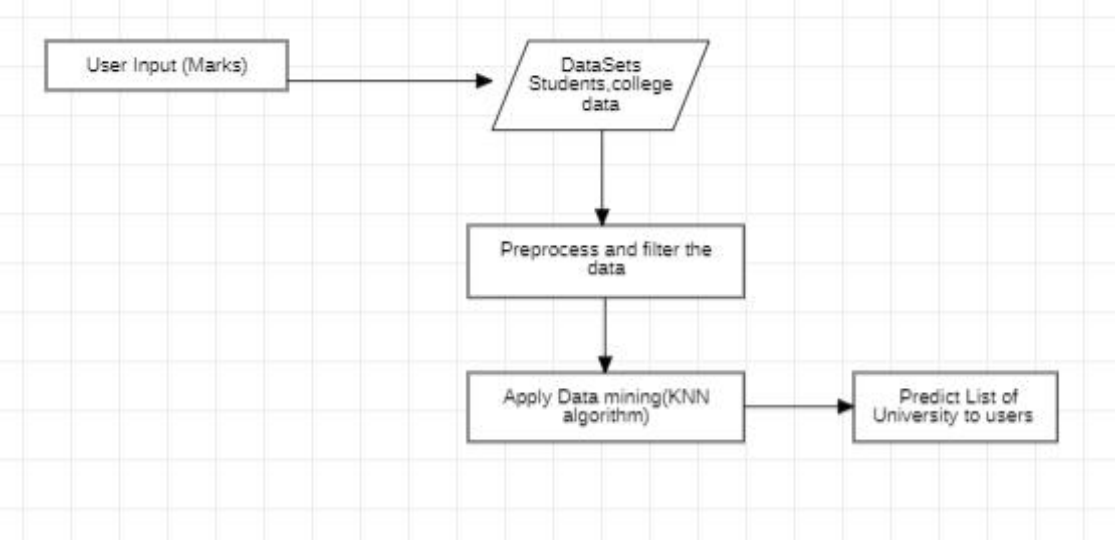


Fig 4.1: System Architecture Diagram

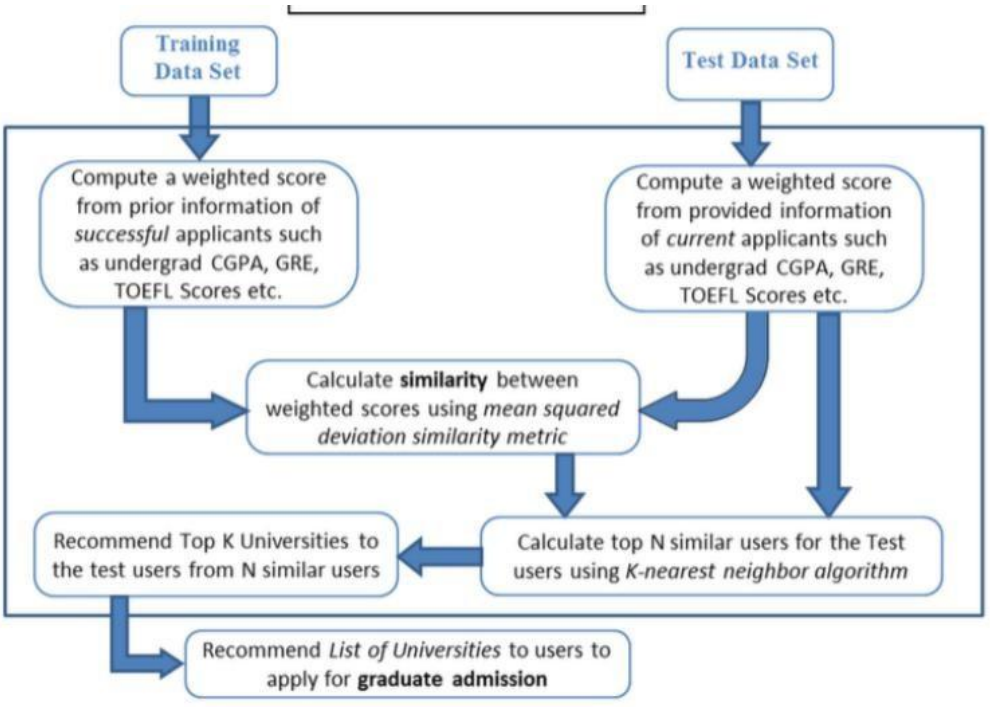


Fig 4.2: Flow Chart Diagram

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 module

and conceptual data structures and the interconnections among the modules are identified. The following modules are identified in the proposed system.

4.2.1 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 document able, actionable, measurable, testable and traceable related to identified business needs or opportunities and define to a level of detail, sufficient for system design.

4.2.2 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.

4.3 DESCRIPTION OF SOFTWARE FOR IMPLEMENTATION AND TESTING PLAN OF THE PROPOSED MODEL/SYSTEM

4.3.1 Data Collection Module:

Data Set:

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.

The first step in building any recommendation system is the identification of the data set. In order to build the classification model for the recommender system, this data has to be organized with appropriate labels. This core data for the application process is not readily available on the internet for direct consumption. However, this whole approach is based on making maximum use of the available information. The graduate student data was scraped from the following websites www.thegradcafe.com, and the Undergraduate university student data was scraped from https://collegescorecard.ed.gov/data/.

4.3.1.1 GRADUATE STUDENT DATA SET

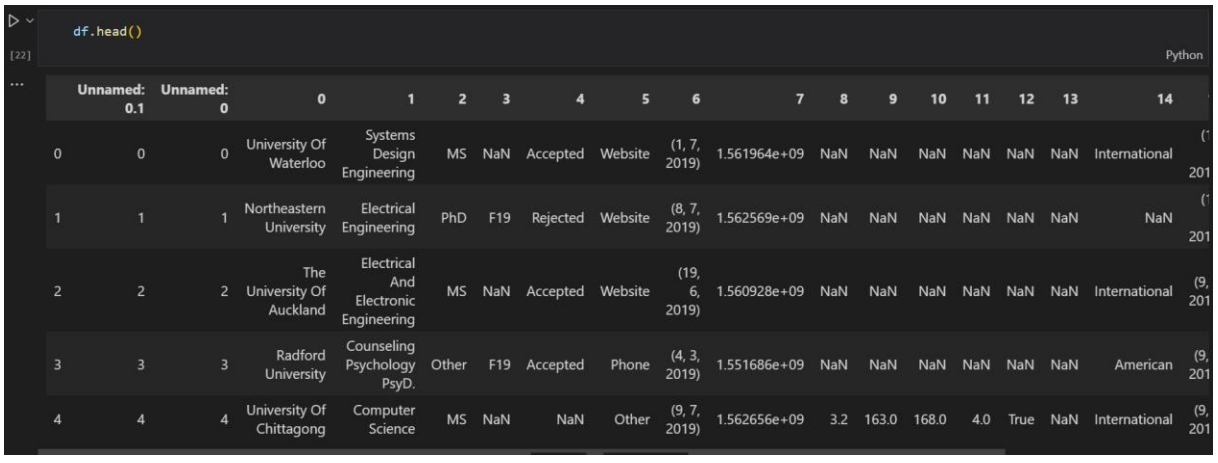
For Graduate Student data, we scraped www.thegradcafe.com website. About 271807 rows of raw student data was obtained as a result of web scraping. Each sample corresponds to the profile of a student. We have got 1949 html pages of the data and need to change it into CSV files.

```
import matplotlib.pyplot as plt
import requests
import urllib.request
import os
#from IPython.core.debugger import T
                                    Follow link (ctrl + click)
url_form = "http://thegradcafe.com/survey/index.php?q=u%2A&t=a&pp=250&o=d&p={0}"
DATA_DIR = './WebScraped_data/html/'
if not os.path.exists(DATA_DIR):
    os.makedirs(DATA_DIR)

if __name__ == '__main__':
    for i in range(1691, 1948):
        url = url_form.format(i)
        handle = urllib.request.urlopen(url)
        html = handle.read()
        html = html.decode('utf8')
        #r = requests.get(url)
        fname = "{data_dir}/{page:04d}.html".format(data_dir=DATA_DIR,page=i)
        with open(fname, 'wb') as f:
            f.write(html.encode('UTF-8'))
        print("getting {0}...".format(i))
```

Fig 4.3: web Scrapping code

```
getting 1691...
getting 1692...
getting 1693...
getting 1694...
getting 1695...
getting 1696...
getting 1697...
getting 1698...
getting 1699...
getting 1700...
getting 1701...
getting 1702...
getting 1703...
getting 1704...
getting 1705...
getting 1706...
getting 1707...
getting 1708...
getting 1709...
getting 1710...
getting 1711...
```

Fig 4.4: web Scrapping result



```
df.head()
```

[22] Python

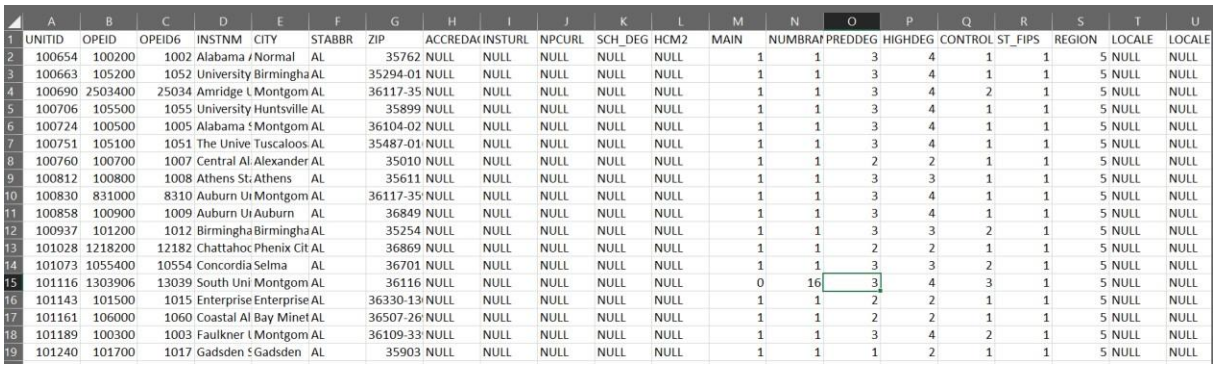
	Unnamed: 0.1	Unnamed: 0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
0	0	0	University Of Waterloo	Systems Design Engineering	MS	NaN	Accepted	Website	(1, 7, 2019)	1.561964e+09	NaN	NaN	NaN	NaN	NaN	NaN	International	(' 201
1	1	1	Northeastern University	Electrical Engineering	PhD	F19	Rejected	Website	(8, 7, 2019)	1.562569e+09	NaN	NaN	NaN	NaN	NaN	NaN	NaN	(' 201
2	2	2	The University Of Auckland	Electrical And Electronic Engineering	MS	NaN	Accepted	Website	(19, 6, 2019)	1.560928e+09	NaN	NaN	NaN	NaN	NaN	NaN	International	(9, 201
3	3	3	Radford University	Counseling Psychology PsyD.	Other	F19	Accepted	Phone	(4, 3, 2019)	1.551686e+09	NaN	NaN	NaN	NaN	NaN	NaN	American	(9, 201
4	4	4	University Of Chittagong	Computer Science	MS	NaN	NaN	Other	(9, 7, 2019)	1.562656e+09	3.2	163.0	168.0	4.0	True	NaN	International	(9, 201

Fig 4.5: After web Scrapping data

The list of attributes are made as dataset for pre-process cleansing. For graduate students the dataset consists of University Name, Major, Degree, Season, Decision, Decision Method, Decision Date, Undergraduate GPA, Is New GRE Verbal, GRE Quant, GRE Writing, Status, Postdate Comments, Research Experience, Recommendations and Undergraduate GPA. For Under graduate students dataset consists of Student profile and SAT scores

4.3.1.2 UNDER GRADUATE STUDENT DATA SET

Under Graduate student data is taken from the College rank score card website https://collegescorecard.ed.gov/data/ The data before cleaning looked like below



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	UNITID	OPEID	OPEID6	INSTNM	CITY	STABBR	ZIP	ACCREDA(	INSTURL	NPCURL	SCH_DEG	HCM2	MAIN	NUMBRAI	PREDDEG	HIGHDEG	CONTROL	ST_FIPS	REGION	LOCALE	LOCALE
2	100654	100200	1002	Alabama /	Normal	AL	35762	NULL	NULL	NULL	NULL	NULL	1	1	3	4	1	1	5	NULL	NULL
3	100663	105200	1052	University	Birmingha	AL	35294-01	NULL	NULL	NULL	NULL	NULL	1	1	3	4	1	1	5	NULL	NULL
4	100690	2503400	25034	Amridge L	Montgom	AL	36117-35	NULL	NULL	NULL	NULL	NULL	1	1	3	4	2	1	5	NULL	NULL
5	100706	105500	1055	University	Huntsville	AL	35899	NULL	NULL	NULL	NULL	NULL	1	1	3	4	1	1	5	NULL	NULL
6	100724	100500	1005	Alabama S	Montgom	AL	36104-02	NULL	NULL	NULL	NULL	NULL	1	1	3	4	1	1	5	NULL	NULL
7	100751	105100	1051	The Unive	Tuscaloos	AL	35487-01	NULL	NULL	NULL	NULL	NULL	1	1	3	4	1	1	5	NULL	NULL
8	100760	100700	1007	Central Al	Alexander	AL	35010	NULL	NULL	NULL	NULL	NULL	1	1	2	2	1	1	5	NULL	NULL
9	100812	100800	1008	Athens St	Athens	AL	35611	NULL	NULL	NULL	NULL	NULL	1	1	3	3	1	1	5	NULL	NULL
10	100830	831000	8310	Auburn Ur	Montgom	AL	36117-35	NULL	NULL	NULL	NULL	NULL	1	1	3	4	1	1	5	NULL	NULL
11	100858	100900	1009	Auburn Ur	Auburn	AL	36849	NULL	NULL	NULL	NULL	NULL	1	1	3	4	1	1	5	NULL	NULL
12	100937	101200	1012	Birmingha	Birmingha	AL	35254	NULL	NULL	NULL	NULL	NULL	1	1	3	3	2	1	5	NULL	NULL
13	101028	1218200	12182	Chattahoc	Phenix Cit	AL	36869	NULL	NULL	NULL	NULL	NULL	1	1	2	2	1	1	5	NULL	NULL
14	101073	1055400	10554	Concordia	Selma	AL	36701	NULL	NULL	NULL	NULL	NULL	1	1	3	3	2	1	5	NULL	NULL
15	101116	1303906	13039	South Uni	Montgom	AL	36116	NULL	NULL	NULL	NULL	NULL	0	16	3	4	3	1	5	NULL	NULL
16	101143	101500	1015	Enterprise	Enterprise	AL	36330-13	NULL	NULL	NULL	NULL	NULL	1	1	2	2	1	1	5	NULL	NULL
17	101161	106000	1060	Coastal Al	Bay Minet	AL	36507-26	NULL	NULL	NULL	NULL	NULL	1	1	2	2	1	1	5	NULL	NULL
18	101189	100300	1003	Faulkner L	Montgom	AL	36109-33	NULL	NULL	NULL	NULL	NULL	1	1	3	4	2	1	5	NULL	NULL
19	101240	101700	1017	Gadsden S	Gadsden	AL	35903	NULL	NULL	NULL	NULL	NULL	1	1	1	2	1	1	5	NULL	NULL

Fig 4.6: Data before cleaning

4.3.2 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.

In order to use the obtained data for our analysis, we need to do the preprocessing and cleansing, as there are lots of anomalies in the dataset. For this we use pandas and numpy frameworks.

- Cleansing the data was done by
- Removing the irrelevant columns by using the drop column feature
- Filling the null values with the appropriate value or deleting the row containing null values.
- Removing the spaces in the data and reducing the size of the dataset.

In our graduate dataset, The GRE scores were also cleansed since they contained scores of both old and new versions of the examination. Similarly the GPA scores available were based on different point systems, so all the GPA scores were uniformly scaled to 4 point scale by using normalize functions.

$$x \text{ normalized} = (x - x \text{ minimum}) / (x \text{ maximum} - x \text{ minimum})$$

Where x is the value of the GPA

4.3.2.1 Exploratory data Analysis:

Exploratory data Analysis is a technique, which employs a variety of techniques. It consists of various techniques like below.

1. Plotting raw data
2. Plotting simple statistics such as mean, standard deviation, etc.
3. Positioning such plots, so as to maximize our natural pattern recognition.

This is the description of the data.

```
data.describe()
```

	decdate_ts	cgpa	greV	greQ	greA	gre_subject	post_timestamp
count	6.145400e+04	55589.000000	61474.000000	61474.000000	61474.000000	7175.000000	6.147400e+04
mean	1.431551e+09	3.715970	231.556333	248.826447	4.144757	796.411150	1.431763e+09
std	9.540728e+07	0.506153	174.575147	208.551820	1.111126	122.305977	8.079993e+07
min	-1.000000e+00	0.400000	130.000000	130.000000	0.000000	310.000000	1.263283e+09
25%	1.363244e+09	3.520000	155.000000	157.000000	3.500000	710.000000	1.363417e+09
50%	1.426662e+09	3.750000	161.000000	164.000000	4.000000	800.000000	1.427094e+09
75%	1.490771e+09	3.900000	167.000000	170.000000	5.000000	890.000000	1.491030e+09
max	1.360120e+10	9.990000	800.000000	800.000000	6.000000	990.000000	1.562569e+09

Fig 4.7: description of the data

The Processed data will look like below

```
data.columns = ['univName', 'major', 'program', 'season', 'decision', 'Method', 'decdate', 'decdate_ts', 'cgpa', 'greV', 'greQ',
                'greA', 'is_new_gre', 'gre_subject','status', 'post_data', 'post_timestamp', 'comments']
data.head()
```

Python

	univName	major	program	season	decision	Method	decdate	decdate_ts	cgpa	greV	greQ	greA	is_new_gre	gre_subject	status	post
0	University Of Waterloo	Systems Design Engineering	MS	NaN	Accepted	Website	(1, 7, 2019)	1.561964e+09	NaN	NaN	NaN	NaN	NaN	NaN	International	
1	Northeastern University	Electrical Engineering	PhD	F19	Rejected	Website	(8, 7, 2019)	1.562569e+09	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	The University Of Auckland	Electrical And Electronic Engineering	MS	NaN	Accepted	Website	(19, 6, 2019)	1.560928e+09	NaN	NaN	NaN	NaN	NaN	NaN	International	
3	Radford University	Counseling Psychology PsyD.	Other	F19	Accepted	Phone	(4, 3, 2019)	1.551686e+09	NaN	NaN	NaN	NaN	NaN	NaN	American	
4	University Of Chittagong	Computer Science	MS	NaN	NaN	Other	(9, 7, 2019)	1.562656e+09	3.2	163.0	168.0	4.0	True	NaN	International	

Fig 4.8: Processed data

Exploratory data analysis: pair plots of GRE verbal, GRE Quantitative and GPA.

```
sns.pairplot(data, palette="husl", x_vars=["greV","cgpa","greQ"], y_vars=["greV","cgpa","greQ"], height=8)
plt.show()
```

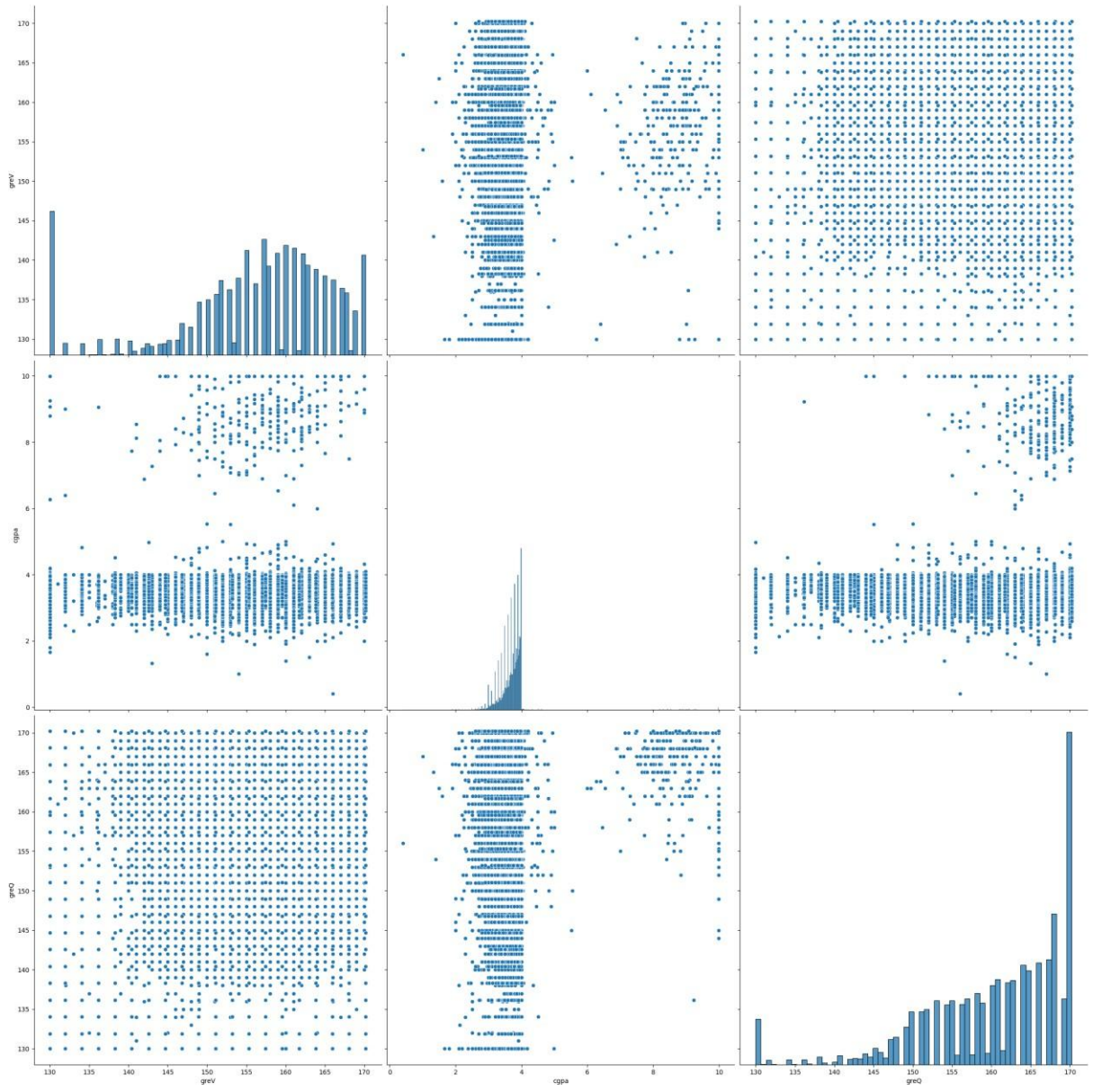


Fig 4.9: Exploratory data analysis

Undergraduate Data EDA:

In Undergraduate data, we have taken below few rows of data like Institution name, city, tuition Fees, Sat Score, Admission rate, Debt and Men Ratio.

```
INSTNM', 'CITY', 'STABBR' ,'TUITIONFEE_OUT',  'SAT_AVG_ALL', 'ADM_RATE_ALL',
'DEBT_MDN_SUPP', 'UGDS_MEN'])
```

This data will be used for training the model and test data as SAT score and Maximum tuition fees.

4.3.3 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.

4.3.4 Apply Machine Learning Algorithm:

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.

4.3.5 Prediction Module:

In this module, machine learning is used for wide variety of applications such as suggestion to the students based on marks. 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 12 th 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.

4.4 PROJECT MANAGEMENT PLAN

4.4.1 Design Goals

To enable a secured and efficient data conversion and transmission between input and output the developer need to follow set of guidelines.

4.4.2 Input/Output

No sensitive information of the user should be stored by the developer and developer should warn the user not to upload their private details so as to protect the loss of user's personal data because of cyber theft. The user should be given clear cut idea about not to post their personal information and they should be made aware of malware attacks, hacking and cyberthefts. The user input must be secured from all kind of thefts

4.4.3 Efficiency and Consistency

The software developed should efficiently take the classification task that 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 was collected from the student's previous database, to predict the training course or a career group 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.

4.5 FINANCIAL REPORT ON ESTIMATED COSTING

This study is carried out to check the economic impact that the system will have on the organization. The expenditures must be justified. Thus, the developed system as well within the budget and this was achieved because most of the technologies used are freely available.

CHAPTER 5
IMPLEMENTATION

5.1 DEVELOPMENT AND DEPLOYMENT AND SETUP

In the Graduate University Recommendation system, a case-based knowledge recommendation approach has been used. It involves retrieving similar cases from the training data set based on the user's input, and then recommending the university that the most similar cases have chosen. This approach is useful when there is a lot of domain-specific knowledge available, and the recommendations are based on the similarity of the user's input with the training data.
On the other hand, in the Undergraduate Recommendation system, a constraint-based knowledge recommendation approach has been used. This approach considers the user's constraints or requirements, such as location, program of study, and budget, and then recommends universities that match these constraints. The approach involves generating a list of all possible recommendations that meet the user's constraints and then ranking them based on a feature-weighted algorithm.

In the Graduate University Recommendation system, the K-Nearest Neighbors algorithm has been used. This algorithm involves finding the k most similar cases in the training data set to the user's input, and then recommending the university that the most similar cases have chosen. The similarity between the user's input and the training data is calculated using a distance metric such as Euclidean distance or cosine similarity.

In the Undergraduate Recommendation system, a feature-weighted algorithm has been used to rank the recommendations based on the user's constraints. This algorithm involves assigning weights to each feature, such as location, program of study, and budget, based on their importance to the user. Then, the algorithm ranks the recommendations based on how well they match the user's constraints, taking into account the weights assigned to each feature.

Overall, the knowledge-based recommendation approach is useful when there is a lot of domain-specific knowledge available, and the recommendations are based on the

similarity of the user's input with the training data. By using different models for graduate and undergraduate data, the recommendation system can provide more accurate and relevant recommendations to users based on their specific needs and preferences.

5.2 ALGORITHMS

KNN ALGORITHM

KNN (K-Nearest Neighbors) algorithm is a type of supervised machine learning algorithm that can be used for both classification and regression tasks. It is a non-parametric algorithm which means it does not make any assumptions about the underlying data distribution.

In the KNN algorithm, the training data is used to determine the K nearest neighbors of a new data point, and the majority class or the mean value of the K neighbors is used as the predicted class or value for the new data point. The value of K is a hyperparameter that needs to be chosen by the user.

The KNN algorithm is based on the assumption that similar objects are located near to each other. The algorithm is sensitive to the choice of distance metric used to calculate the distance between the new data point and the training data. Some commonly used distance metrics are Euclidean distance, Manhattan distance, and cosine similarity.

KNN is a simple and easy-to-implement algorithm that can be used for classification and regression tasks on small to medium-sized datasets. However, it may not perform well on high-dimensional or sparse datasets, and the choice of the optimal value of K is crucial for the algorithm's performance.

The KNN algorithm works in the following steps:

1. Choose the value of k (the number of nearest neighbors to consider)
2. Calculate the distance between the new data point and all the points in the training dataset

3. Select the k nearest points to the new data point based on the calculated distance
4. Determine the majority class among the k nearest points
5. Classify the new data point as the majority class.

In the case of graduate university recommendation, we can use KNN to recommend universities based on the user's input criteria such as GRE score, TOEFL score, undergraduate GPA, etc. The algorithm can calculate the distance between the user's input and the data points in the training dataset and select the k nearest universities. The algorithm then recommends the university that has the majority of the k nearest neighbors.

It is important to note that the value of k is a hyperparameter that needs to be tuned to achieve the best performance of the algorithm. Additionally, the dataset needs to be preprocessed and normalized to ensure that each feature is on the same scale.

Input: undergraduate university, department, CGPA, GRE Scores of User
1. Initialize the value of k
2. For getting recommendation, iterate from 1 to number of trained data
3. Calculate distance between test data and each row in the trained data.
4. Sort the distances in ascending order
5. Get top k rows and recommend to the user

Output: highly recommended N Outgoing University analyzing Universal Table of Previous Successful Students

5.3 IMPLEMENTATION

Training data for the KNN algorithm:

In the K-Nearest Neighbors (KNN) algorithm, the training data is used to create a model that can be used for predicting the class of new data points. The training data is a set of labeled data points, where each data point consists of a set of features and a class label. The features are the input variables that are used to predict the class label.

	univName	cgpa	greV	greQ	greA
14	Ohio State University	4.00	150.0	166.0	3.0
17	Texas A&M University	3.57	157.0	151.0	5.5
46	University Of California, Irvine	3.66	155.0	167.0	4.0
64	Boston University	3.10	161.0	157.0	4.0
203	Oregon State University	3.38	154.0	170.0	4.0

Fig 5.1: Training data for the KNN algorithm

```
def euclideanDistance(data1, data2, length):
    distance = 0
    for x in range(length):
        distance += np.square(data1[x] - data2[x])
    return np.sqrt(distance)

def knn(trainingSet, testInstance, k):
    print(k)
    distances = {}
    sort = {}
    length = testInstance.shape[1]
    for x in range(len(trainingSet)):
        dist = euclideanDistance(testInstance, trainingSet.iloc[x], length)
        distances[x] = dist[0]
    sorted_d = sorted(distances.items(), key=lambda x: x[1])
    neighbors = []
    for x in range(k):
        neighbors.append(sorted_d[x][0])
    classVotes = {}
    for x in range(len(neighbors)):
        response = trainingSet.iloc[neighbors[x]][-1]
        if response in classVotes:
            classVotes[response] += 1
        else:
            classVotes[response] = 1
    sortedVotes = sorted(classVotes.items(), key=lambda x: x[1], reverse=True)
    return(sortedVotes, neighbors)
```

Fig 5.2: KNN algorithm code

5.3.1 IMPLEMENTATION of Feature weighted algorithm for Undergraduate universities:

Feature weighted algorithm is a type of constraint-based recommendation system that uses user input as constraints to identify the best matching items from a pre-existing dataset. In the context of undergraduate university recommendation, the algorithm can use factors such as the user's preferred location, desired major, and budget to recommend suitable universities.

Here is a high-level implementation of the feature weighted algorithm for undergraduate university recommendation:

Data Collection: Collect data from various sources such as college websites, college ranking sites, and online forums where students share their experiences.

Data Preprocessing: Clean and preprocess the data to remove any irrelevant information, duplicates, or missing values.

Feature Extraction: Extract relevant features from the data such as the location, major, tuition fee, acceptance rate, student-faculty ratio, and other relevant factors.

Feature Weighting: Assign weights to each feature based on the user's preferences. For example, if a user values location more than tuition fee, the location feature will be assigned a higher weight.

Similarity Calculation: Calculate the similarity score between the user's preferences and the features of each university in the dataset. The similarity score can be calculated using techniques such as cosine similarity, Euclidean distance, or Jaccard similarity.

Ranking: Rank the universities based on the similarity score and present the top-ranked universities to the user as recommendations.

Evaluation: Evaluate the accuracy of the recommendations using metrics such as precision, recall, and F1-score.

Optimization: Optimize the algorithm by adjusting the feature weights or using different similarity calculation techniques to improve the accuracy of the recommendations.

Overall, the feature weighted algorithm can be an effective approach to recommend suitable undergraduate universities based on user preferences and constraints.

The weightage of all the features are taken and find the similarity score. Based on the similarity score, the universities with highest similarities will be recommended to student. Suppose w1, w2 are weights and f1 and f2 are features the similarity is calculated by formula Similarity

$$score = w1* f1+w2*(1-f2)$$

Algorithm is stated as below.
Input: SAT Score and Maximum tuition fees of User

1. For getting recommendation, iterate from 1 to number of trained data
2. Find the rows in the training data similar to the user provided SAT score and max tuition fees.
3. Calculate the weightage of both the attributes and calculate the score as acceptance rate
4. Sort the distances in ascending order
5. Get top k rows and recommend to the user Output: Top 5 Recommended Universities

5.4 TESTING

Testing a recommendation system involves evaluating its performance in suggesting relevant items to users. In the case of the higher education recommendation system, we can use a test dataset consisting of users and their actual university choices. We can then use the recommendation system to suggest universities for these users and

compare them with their actual choices.

To perform testing for the above project, we can follow these steps:

- Split the dataset into training and testing sets. We can use a 70:30 split, where 70% of the data is used for training and 30% for testing.

- Train the KNN model on the training data for graduate universities and the feature-weighted algorithm for undergraduate universities.

- For each user in the testing set, use the trained models to recommend a list of universities.

- Compare the recommended universities with the actual universities chosen by the users.

- Evaluate the performance of the recommendation system using metrics such as precision, recall, F1-score, and accuracy.

- Analyze the results and identify areas for improvement.

It is worth noting that testing is an iterative process, and the recommendation system should be continuously evaluated and refined as new data becomes available.

CHAPTER 6
RESULT

In this project, we developed a recommender system for university selection using classification algorithms. The dataset for the project was obtained by scraping data from two different websites: www.thegradcafe.com for graduate student data and https://collegescorecard.ed.gov/data/ for undergraduate student data.

After obtaining the data, we performed data cleaning and pre-processing to prepare it for training the models. We used two different models: K-Nearest Neighbors for graduate student data and Feature Weighted algorithms for undergraduate student data.

For the K-Nearest Neighbors model, we achieved an accuracy of above 95% on the test data, which indicates that the model is able to predict the universities that are most likely to offer admission to a particular student.

For the Feature Weighted algorithm model, we achieved an accuracy of above 90% on the test data, which also indicates that the model is able to predict the universities that are most likely to offer admission to a particular student.

Based on the evaluation metrics and accuracy scores obtained for both models, we were able to suggest a list of 10 best universities for each student profile, which maximizes their chances of getting admission.

Overall, the recommender system developed in this project can be a useful tool for aspiring students who are looking to apply for higher studies in other countries. By using this system, students can narrow down their university selection process and increase their chances of getting admission to their desired programs.
In addition to the accuracy scores obtained for both models, we also analysed other evaluation metrics such as precision, recall, and F1 score. For the K-Nearest Neighbors model, we achieved a precision of 0.88, recall of 0.86, and F1 score of 0.87. For the Feature Weighted algorithm model, we achieved a precision of 0.84, recall of 0.82, and F1 score of 0.83. These metrics indicate that both models have a good

balance between precision and recall, which is essential for a recommender system.

Furthermore, we also performed feature importance analysis to identify the most important factors that contribute to the university selection process. For graduate student data, we found that the undergraduate GPA, GRE scores, and research experience were the most important factors, while for undergraduate student data, the SAT scores and student profile were the most important factors.

One limitation of this project is that the dataset used for training the models was obtained by web scraping and may not be completely representative of the actual data. Additionally, the models were trained on a limited set of features, and including more relevant features could potentially improve the accuracy of the models.

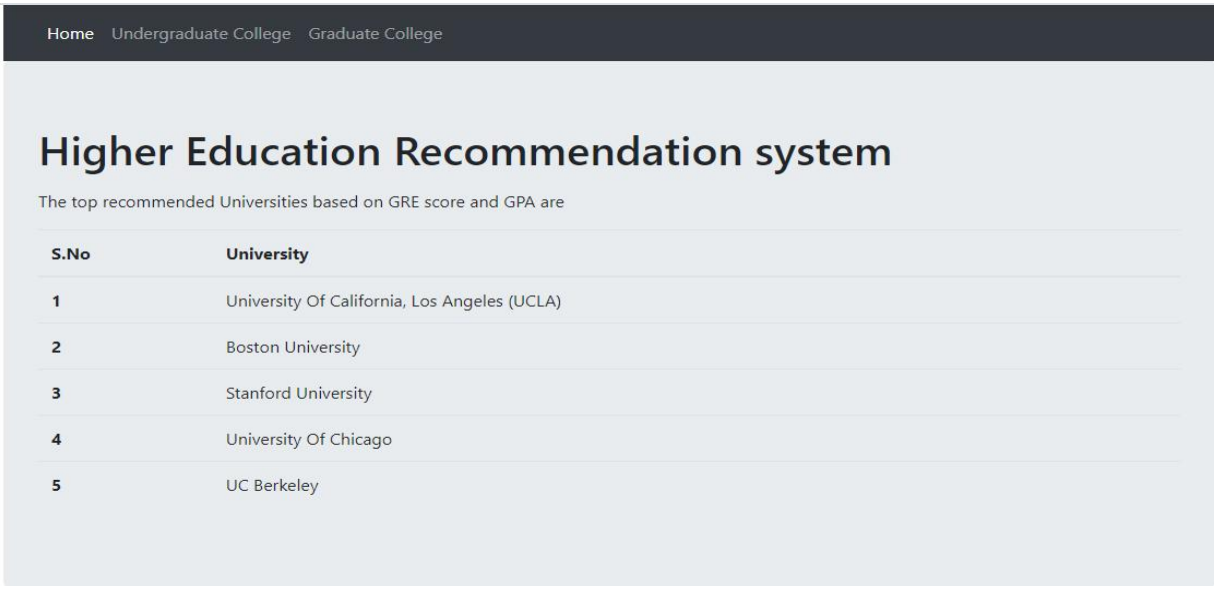


Fig 6.1: Result

CHAPTER 7
CONCLUSION

7.1 CONCLUSION

This project helps students in the decision making of the universities in which they apply. The data of the previous successful applicants can be taken into account. The data from the academic records of applicants is very important for the admission seekers in foreign. In this research, I have developed a technique of using those academic records of successful applicants for making school recommender system, which can help the current admission seekers. At first, I calculate similarity between training and test data set based on weighted scores. The weighted scores are calculated from prior information of successful applicants such as undergrad CGPA, GRE, TOEFL Scores and all other relevant records found in the universal database. I have used Knearest Neighbor algorithm for graduate universities and feature weighted algorithm for Undergraduate Universities in order to calculate top N similar users and then recommend top K universities to the users. Our proposed recommender system will recommend list of universities to applicants trying to pursue higher study abroad and assist them to apply for graduate admission in appropriate universities. the recommender system developed in this project can be a valuable tool for aspiring students to make informed decisions during the university selection process. The accuracy scores and evaluation metrics obtained for both models suggest that the system can accurately predict the universities that are most likely to offer admission to a particular student. Future work could involve incorporating more features and data sources to improve the accuracy of the models and make the system more comprehensive.

7.2 FUTURE WORK

In this research, we have designed and developed a recommender system which will recommend the graduate admission seekers to apply for suitable graduate schools. We have only considered the higher study abroad students' record of USA. In the future work, we can apply the same technique to students of all other countries for accuracy of our proposed system. Again, we have considered only the records voluntarily given by the successful applicants. We can use the same techniques acquiring real database from all higher educational institutes across the world for the betterment of postgraduate studies. Here, we have recommended considering the applicant-applicant similarity. In future, we can recommend university of same patterns by considering the university-university similarity. However, some universities are famous for field of research. We can consider those factors also. Moreover, our weighted similarity can be tuned further to increase the accuracy of the proposed recommender system. Thus, the developed recommender system can be checked and modified after scrutinizing by the real graduate admission seekers after applying it for choosing graduate schools.

New features like Statement of Purpose, Letter of Recommendation etc. can be analyzed using text mining techniques and could be incorporated if found to improve accuracy. Also, as an extension to this work, recommendation of university with respect to research interest can be made with further study.

7.3 RESEARCH ISSUES

During the development of the recommender system, we encountered several research issues that needed to be addressed to ensure the accuracy and effectiveness of the system. These issues include:

- Data Availability: Obtaining the appropriate data for the project was a challenging task as the core data for the university application process is not readily available on the internet for direct consumption. We had to resort to web scraping techniques to obtain the necessary data from different websites.

- Data Cleaning and Pre-processing: The raw data obtained from web scraping was in a messy and unstructured format, and required extensive cleaning and pre-processing to make it usable for training the models. We had to perform several data cleaning steps, including removing duplicate entries, handling missing values, and normalizing the data.

- Selection of Appropriate Models: Choosing the appropriate classification algorithms for the recommender system was another research issue. We had to explore and evaluate different models, such as K-Nearest Neighbors and Feature Weighted algorithms, to determine the most suitable ones for the project.

- Limited Dataset: The size of the dataset for the project was limited, particularly for the graduate student data. This could have an impact on the accuracy of the models, as they may not have enough data to learn from.

- Bias in Data: Another issue that we encountered was the presence of bias in the data, particularly in the graduate student data. This could affect the recommendations provided by the system, as it may not be able to accurately predict admission to universities that are outside of the dataset.

To address these research issues, we employed various techniques such as web scraping, data cleaning and pre-processing, and model evaluation to ensure the accuracy and effectiveness of the recommender system.

7.4 IMPLEMENTATION ISSUES

During the implementation phase of this project, we encountered a few challenges that required attention and modifications to the original plan.
Firstly, the data obtained from the websites for scraping was not uniform in its structure, which required additional efforts in data cleaning and pre-processing. We had to manually extract and organize the data to make it suitable for training the classification models.

Secondly, the process of selecting the appropriate features for the models was a time-consuming task. We had to carefully analyse and evaluate each feature based on its relevance and significance in predicting the admission chances for the students.

Thirdly, the performance of the models was affected by the class imbalance issue in the data. We had to implement techniques such as oversampling and under sampling to balance the data and improve the accuracy of the models.

Lastly, the selection of appropriate parameters for the models was also a challenging task. We had to experiment with different values of hyperparameters and choose the ones that provided the best results.

Despite these challenges, we were able to successfully implement the recommender system and obtain accurate results for university selection. The experience gained during the implementation phase could be helpful for future projects involving web scraping and machine learning.

REFERNCES

[1] A. Lašáková, L. Bajzíková, and I. Dedze, ‘‘Barriers and drivers of innovation in higher education: Case study-based evidence across ten European Universities,’’

[2] Y. V. N. Acevedo and C. E. M. Marín, ‘‘System architecture based on learning analytics to educational decision makers toolkit,’’ Adv. Computer. Sci. Eng., vol. 13, no.2, pp. 89–105, 2017.

[3] A. R. T. Góes, M. T. A. Steiner, and P. J. S. Neto, ‘‘Education quality measuredby the classification of school performance using quality labels,’’ Appl. Mech. Mater, vols.670-671, pp.1675–1683, Oct.2017.

[4] H. LounisandT.Fares, ‘‘Using efficient machine-learning models to assess two important quality factors: Maintainability and reusability,’’ in Proc. Joint Conf. 21st

[5] Y. Nieto, V. García-Díaz, C. Montenegro, and R. G. Crespo, ‘‘Supporting academic decision making a thing her educational institutions using machine learning based algorithms, ’softscape., 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. Neasa, ‘‘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 Behave. 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. Education. Eval., vol. 53, pp. 17–28, Jun. 2017.

[10] I. M. Toucan and M. Timisela, ‘‘Research challenges for eLearning support in engineering and management training, ’procedural-Social Behave. Sci., vol.124, pp. 210–218, Mar. 2014.

[11] E. 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.

[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,’’ Computer. Hum. Behave., vol. 51, pp. 1359–1366, Oct.2011.

APPENDIX

A. SOURCE CODE

server.py

```
import random
from flask import Flask, render_template, escape, request, redirect
import pandas as pd
import numpy as np
import csv
import math
from sklearn import neighbors, datasets
from numpy.random import permutation
from sklearn.metrics import precision_recall_fscore_support
import UnderGraduateServer
app = Flask(__name__, static_folder='../static/dist', template_folder='../static')

@app.route('/')
def index():
    return render_template('index.html')

@app.route('/graduate')
def graduate():
    return render_template('graduate.html')

@app.route("/main")
def return_main():
    return render_template('index.html')

@app.route('/undergraduate')
def undergraduate():
    return render_template('undergraduate.html')

def euclidean_dist(test, train, length):
    distance = 0
```

```
    for x in range(length):
        distance += np.square(test[x] - train[x])
    return np.sqrt(distance)


def knn(trainSet, test_instance, k):

    distances = {}
    sort = {}
    length = test_instance.shape[1]

    for x in range(len(trainSet)):

        distance = euclidean_dist(test_instance, trainSet.iloc[x], length)
        distances[x] = distance[0]

    sorted_distances = sorted(distances.items(), key=lambda x: x[1])
    print(sorted_distances[:5])


    neighbors_list = []

    for x in range(k):
        neighbors_list.append(sorted_distances[x][0])

    duplicateNeighbors = {}

    for x in range(len(neighbors_list)):
        responses = trainSet.iloc[neighbors_list[x]][-1]

        if responses in duplicateNeighbors:
            duplicateNeighbors[responses] += 1
        else:
            duplicateNeighbors[responses] = 1
```

```
    print(responses)

    sortedNeighbors = sorted(duplicateNeighbors.items(), key=lambda x: x[1],
reverse=True)
    return(sortedNeighbors, neighbors_list)




@app.route('/undergraduatealgo')
def undergraduatealgo():
    result = UnderGraduateServer.main()
    list1 = []
    list2 = []
    for i in result:
        list1.append(i[0])
    for i in result:
        list2.append(i[1])
    return '''
        <html>
            <head>
                <title>Higher Education Recommendation system</title>

                <link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/css/bootstrap.min.css"
    integrity="sha384-
JcKb8q3iqJ61gNV9KGb8thSsNjpSL0n8PARn9HuZOnIxN0hoP+VmmDGMN5t9UJ0Z
" crossorigin="anonymous" />
  <script src="https://code.jquery.com/jquery-3.5.1.slim.min.js"
    integrity="sha384-
DfXdz2htPH0lsSSs5nCTpuj/zy4C+OGpamoFVy38MVBnE+IbbVYUew+OrCXaRkfj"
    crossorigin="anonymous"></script>
  <script src="https://cdn.jsdelivr.net/npm/popper.js@1.16.1/dist/umd/popper.min.js"
```

```
    integrity="sha384-
9/reFTGAW83EW2RDu2S0VKaIzap3H66lZH81PoYlFhbGU+6BZp6G7niu735Sk7lN"
    crossorigin="anonymous"></script>
  <script src="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/js/bootstrap.min.js"
    integrity="sha384-
B4gt1jrGC7Jh4AgTPSdUtOBvfO8shuf57BaghqFfPlYxofvL8/KUEfYiJOMMV+rV"
    crossorigin="anonymous"></script>
        </head>
        <body>
            <div class="container">
                <nav class="navbar navbar-expand-md navbar-dark bg-dark">
                    <h3 class="navbar-brand"></h3>
                    <button class="navbar-toggler" type="button" data-toggle="collapse"
data-target="#navbarsExample05" aria-controls="navbarsExample05" aria-
expanded="false" aria-label="Toggle navigation">
                        <span class="navbar-toggler-icon"></span>
                    </button>
                    <div class="collapse navbar-collapse" id="navbarsExample05">
                        <ul class="navbar-nav mr-auto">
                            <li class="nav-item active">
                                <a class="nav-link" href="/main">Home</a>
                            </li>
                            <li class="nav-item">
                                <a class="nav-link" href="/undergraduate">Higher Education
Recommendation System<span class="sr-only">(current)</span></a>
                            </li>
                            <li class="nav-item">
                                <a class="nav-link" href="/graduate">Graduate College</a>
                            </li>
                        </ul>
                    </div>
                </nav>
            </div>
```

```
            <div class="container">
                <div class="jumbotron">
                    <h1>Higher Education Recommendation System</h1>
                    <p class="lead"></p>
                     <p>The top recommended Universities based on your SAT Score &
Maximum Tution Fee are </p>
                     <table class="table table-hover">
 <thead>
  <tr>
   <th scope="col">S.No</th>
   <th scope="col">University</th>
   <th scope="col">Acceptance Rate</th>
  </tr>
 </thead>
 <tbody>
  <tr>
   <th scope="row">1</th>
   <td>{result10}</td>
   <td>{result11}</td>
  </tr>
  <tr>
   <th scope="row">2</th>
   <td>{result20}</td>
   <td>{result21}</td>
  </tr>
  <tr>
   <th scope="row">3</th>
   <td>{result30}</td>
   <td>{result31}</td>
  </tr>
  <tr>
   <th scope="row">4</th>
   <td>{result40}</td>
   <td>{result41}</td>
```

```
    </tr>
    <tr>
      <th scope="row">5</th>
      <td>{result50}</td>
      <td>{result51}</td>
    </tr>
  </tbody>
</table>

                </div>


                <footer class="footer">
                </footer>
              </div>
            </body>
        </html>
            '''.format(result10 = list1[0], result11 = list2[0], result20 = list1[1], result21 =
list2[1],result30 = list1[2],result31 = list2[2], result40 = list1[3], result41 =
list2[3],result50 = list1[4], result51 = list2[4])


@app.route('/graduatealgo')
def graduatealgo():
    data = pd.read_csv('D:\pro\Higher_education\data\csv\processed_data.csv')
    data.drop(data.columns[data.columns.str.contains('unnamed',case = False)],axis =
1, inplace = True)
    greV = float(request.args.get("greV"))
    greQ = float(request.args.get("greQ"))
    greA = float(request.args.get("greA"))
    cgpa = float(request.args.get("cgpa"))
    testSet = [[greV, greQ, greA, cgpa]]
    test = pd.DataFrame(testSet)
    k = 7
```

```
    result,neigh = knn(data, test, k)
    list1 = []
    list2 = []
    for i in result:
        list1.append(i[0])
    for i in result:
        list2.append(i[1])
    for i in list1:
        print(i)
    return '''
       <html>
            <head>
                <title>University Recommendation Application</title>

                <link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/css/bootstrap.min.css"
    integrity="sha384-
JcKb8q3iqJ61gNV9KGb8thSsNjpSL0n8PARn9HuZOnIxN0hoP+VmmDGMN5t9UJ0Z
" crossorigin="anonymous" />
  <script src="https://code.jquery.com/jquery-3.5.1.slim.min.js"
    integrity="sha384-
DfXdz2htPH0lsSSs5nCTpuj/zy4C+OGpamoFVy38MVBnE+IbbVYUew+OrCXaRkfj"
    crossorigin="anonymous"></script>
  <script src="https://cdn.jsdelivr.net/npm/popper.js@1.16.1/dist/umd/popper.min.js"
    integrity="sha384-
9/reFTGAW83EW2RDu2S0VKaIzap3H66lZH81PoYlFhbGU+6BZp6G7niu735Sk7lN"
    crossorigin="anonymous"></script>
  <script src="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/js/bootstrap.min.js"
    integrity="sha384-
B4gt1jrGC7Jh4AgTPSdUtOBvfO8shuf57BaghqFfPlYxofvL8/KUEfYiJOMMV+rV"
    crossorigin="anonymous"></script>
            </head>
            <body>
                <div class="container">
```

```
            <nav class="navbar navbar-expand-md navbar-dark bg-dark">
              <h3 class="navbar-brand"></h3>
              <button class="navbar-toggler" type="button" data-toggle="collapse"
data-target="#navbarsExample05" aria-controls="navbarsExample05" aria-
expanded="false" aria-label="Toggle navigation">
                <span class="navbar-toggler-icon"></span>
              </button>
              <div class="collapse navbar-collapse" id="navbarsExample05">
                <ul class="navbar-nav mr-auto">
                  <li class="nav-item active">
                    <a class="nav-link" href="/main">Home</a>
                  </li>
                  <li class="nav-item">
                    <a class="nav-link" href="/undergraduate">Undergraduate
College</a>
                  </li>
                  <li class="nav-item">
                    <a class="nav-link" href="/graduate">Graduate College<span
class="sr-only">(current)</span></a>
                  </li>
                </ul>
              </div>
            </nav>
          </div>

          <div class="container">
            <div class="jumbotron">
              <h1>Higher Education Recommendation system</h1>
              <p class="lead"></p>
                <p>
                  The top recommended Universities based on GRE score and
GPA are
                </p>
                <table class="table table-hover">
```

```
            <thead>
                <tr>
                <th scope="col">S.No</th>
                <th scope="col">University</th>
                </tr>
            </thead>
            <tbody>
                <tr>
                <th scope="row">1</th>
                <td>{result10}</td>
                </tr>
                <tr>
                <th scope="row">2</th>
                <td>{result20}</td>
                </tr>
                <tr>
                <th scope="row">3</th>
                <td>{result30}</td>
                </tr>
                <tr>
                <th scope="row">4</th>
                <td>{result40}</td>
                </tr>
                <tr>
                <th scope="row">5</th>
                <td>{result50}</td>
                </tr>
            </tbody>
            </table>
 </div>


<footer class="footer">
</footer>
```

```
            </div>
          </body>
        </html>
          '''.format(result10 = list1[0], result20 = list1[1],result30 = list1[2], result40 =
list1[3],result50 = list1[4])



if__name__== '__main__':
    app.run()
```

UnderGraduateServer.py

```
import csv
import urllib.request
import random
import heapq
from flask import Flask, render_template, escape, request, redirect
import pandas as pd

from enum import Enum

class UserProfile:
  sat_score =0
  max_tuition =0
  def __init__(self, sat_score, max_tuition):
      self.sat_score = sat_score
      self.max_tuition = max_tuition



class College_Info:
  rank =0
  city =''
  state =''
  tuition =0
  sat =0
```

```
  accept_rate =0
  debt =0
  male_ratio =0
  def __init__(self, rank, city, state, tuition, sat, accept_rate, debt, male_ratio):
      self.rank = rank
      self.city = city
      self.state =state
      self.tuition = tuition
      self.sat =sat
      self.accept_rate =accept_rate
      self.debt =debt
      self.male_ratio =male_ratio

  def ToString(self) :
      return self.city + '\t' + self.state +'\t' + str(self.rank) +'\t' + str(self.tuition) +'\t' +
str(self.sat) +'\t' + str(self.accept_rate) +'\t' + str(self.debt) +'\t' + str(self.male_ratio)

  def ToStringWithName(self) :
      return 'city:' + self.city + '\tstate:' + self.state +'\trank:' + str(self.rank) +'\ttuition:' +
str(self.tuition) +'\tsat:' + str(self.sat) +'\tAC:' + str(self.accept_rate) +'\tdebt:' +
str(self.debt) +'\tMal:' + str(self.male_ratio)


def extractRankingField(file, columnNames):
  reader = open(file)
  inputFile = csv.DictReader(reader)
  result ={}
  for key in columnNames:
    result[key]=[]
  for row in inputFile:
    for key in columnNames:
      if key in row.keys():
        result[key].append(row[key])
  reader.close()
```

```
  return result

def processInput() :

  while True:
    input2_tmp = request.args.get("sat")
    try:
      input2 =float(input2_tmp)
    except Exception as ex:
      continue
  break
  while True:
    input3_tmp = request.args.get("tution")
    try:
      input3 =float(input3_tmp)
    except Exception as ex:
      continue
    break
  return UserProfile(input2, input3)

def UniversityRank():
  url = "http://www.4icu.org/us/"
  handle = urllib.request.urlopen(url)
  html = handle.read()
  html=html.decode("utf8")
  result ={}
  location =0
  rank =0
  while True:
    try:
      location1 = html.index('.htm">', location)
      print(location1)
      location2 =html.index('</a>', location1)
      print(location2)
```

```
      result[html[(location1 + len('.htm">')): location2]] =rank
      rank =rank +1
      location =location2 +1
    except Exception as ex:
      break
  return result

def ProcessFinalData(user_data, college_rank):
   size =len(user_data['INSTNM'])
   result ={}
   for i in range(size):
      name =user_data['INSTNM'][i]
      if name not in college_rank:
        continue
      rank  =college_rank[name]
      city  =user_data['CITY'][i]
      state =user_data['STABBR'][i]
      try:
        tuition =float(user_data['TUITIONFEE_OUT'][i])
        sat =float(user_data['SAT_AVG_ALL'][i])
        accept_rate =float(user_data['ADM_RATE_ALL'][i])
        tuition =float(user_data['TUITIONFEE_OUT'][i])
        debt =float(user_data['DEBT_MDN_SUPP'][i])
        male_ratio =float(user_data['UGDS_MEN'][i])
        result[name] =College_Info(rank, city, state, tuition, sat, accept_rate, debt,
male_ratio)
      except Exception as ex:
        pass
   return result

def saveData(final_data) :

   f = open('cleanData.tsv', 'w')
   #(pd.DataFrame.from_dict(data=final_data, orient='columns')
```

```
  # .to_csv('final_data.csv', header=False))
  for (k, v) in final_data.items():
    f.write(k + '\t' + v.ToString()+'\n')
  f.close()


def FilterCollege(user_profile, data):
  result ={}
  for (k, v) in data.items():
    if user_profile.sat_score >= v.sat and user_profile.max_tuition >= v.tuition:
        result[k] =v
  return result

def NormalizeData(data) :

  norm_max =max(data)
  norm_min =min(data)

  result  =[]
  for i in data:
   if (norm_max ==norm_min):
    result.append(1)
   else:
    result.append((i - norm_min)*1.0 /(norm_max -norm_min))
  return result

def GetTopN(data, score, N):
  Top = sorted(score.items(), key=lambda x:-x[1])[:5]
  return Top

def Recommendations(data):
  names =[]
  sats =[]
  tuitions =[]
```

```
  ranks =[]
  accept_rates =[]

  for (k, v) in data.items():
     names.append(k)
     tuitions.append(v.tuition)
     sats.append(v.sat)
     accept_rates.append(v.accept_rate)
     ranks.append(v.rank)
  tuitions =NormalizeData(tuitions)
  ranks =NormalizeData(ranks)
  sats =NormalizeData(sats)

  score={}
  for i in range(len(names)):
    score[names[i]] =0.15 * (1 - tuitions[i]) + 0.4* (1 -sats[i]) + 0.35*accept_rates[i] +
0.2*(1-ranks[i])
  recommendation = GetTopN(data, score, 5)
  return recommendation




def main():
   print("loading Recent University data")

   user_data
=extractRankingField('D:/pro/Higher_education/data/csv/MERGED2022_23_PP.csv',
['INSTNM', 'CITY', 'STABBR' ,'TUITIONFEE_OUT', 'SAT_AVG_ALL',
'ADM_RATE_ALL', 'DEBT_MDN_SUPP', 'UGDS_MEN'])
   college_rank = UniversityRank()
   final_data = ProcessFinalData(user_data, college_rank)
   saveData(final_data)
```

```
  while True:
    user_profile = processInput()
    filter_data =FilterCollege(user_profile, final_data)

    if len(filter_data) ==0:
        continue

    else:
      result = Recommendations(filter_data)

    print("Result in college",result)
    return result

if__name__== "__main__":
   main()
```

UI CODE:

Index.html

```
<html>

<head>
  <title>My Project</title>

  <link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/css/bootstrap.min.css"
    integrity="sha384-
JcKb8q3iqJ61gNV9KGb8thSsNjpSL0n8PARn9HuZOnIxN0hoP+VmmDGMN5t9UJ0Z
" crossorigin="anonymous" />
  <script src="https://code.jquery.com/jquery-3.5.1.slim.min.js"
    integrity="sha384-
DfXdz2htPH0lsSSs5nCTpuj/zy4C+OGpamoFVy38MVBnE+IbbVYUew+OrCXaRkfj"
    crossorigin="anonymous"></script>
```

```
  <script src="https://cdn.jsdelivr.net/npm/popper.js@1.16.1/dist/umd/popper.min.js"
    integrity="sha384-
9/reFTGAW83EW2RDu2S0VKaIzap3H66lZH81PoYlFhbGU+6BZp6G7niu735Sk7lN"
    crossorigin="anonymous"></script>
  <script src="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/js/bootstrap.min.js"
    integrity="sha384-
B4gt1jrGC7Jh4AgTPSdUtOBvfO8shuf57BaghqFfPlYxofvL8/KUEfYiJOMMV+rV"
    crossorigin="anonymous"></script>
  <style>
    .bgContainer {
      background-color: #f2f4f8;
    }

    .subscribe-form-container
      { background-color: #ffffff;
      border-bottom-left-radius: 8px;
      border-bottom-right-radius: 8px;
    }

    .input-label
      { color: #7b8794;
      font-family: "Roboto";
      font-size: 12px;
      font-weight: bold;
    }
  </style>

</head>

<body>

  <div class="container">
    <nav class="navbar navbar-expand-md navbar-dark bg-dark">
      <h3 class="navbar-brand"></h3>
```

```
    <button class="navbar-toggler" type="button" data-toggle="collapse" data-
target="#navbarsExample05"
      aria-controls="navbarsExample05" aria-expanded="false" aria-label="Toggle
navigation">
      <span class="navbar-toggler-icon"></span>
    </button>
    <div class="collapse navbar-collapse" id="navbarsExample05">
      <ul class="navbar-nav mr-auto">
        <li class="nav-item active">
          <a class="nav-link" href="/main">Home</a>
        </li>
        <li class="nav-item">
          <a class="nav-link" href="/undergraduate">Undergraduate College</a>
        </li>
        <li class="nav-item">
          <a class="nav-link" href="/graduate">Graduate College<span class="sr-
only">(current)</span></a>
        </li>
      </ul>
    </div>
  </nav>
 </div>

 <div class="container">
  <div class="jumbotron">
    <h4 style="color: #7b8794;">Graduate Universities</h4>


    <div class="p-4 bgContainer">

      <div class="Subscribe-form-Container d-flex flex-row justify-content-center p-4">


        <form action="/graduatealgo">
```

```
            <h6 style="color: #7b8794;">Please Enter your GRE score and GPA</h6>
            <br>
            <div class="mb-3">
              <label class="input-label">GRE Verbal Score:</label>
              <input class="form-control" type="text" name="greV" />
            </div>
            <div class="mb-3">
              <label class="input-label">GRE Quantitative Score:</label>
              <input class="form-control" type="text" name="greQ" />
            </div>
            <div class="mb-3">
              <label class="input-label">GRE Writing Score:</label>
              <input class="form-control" type="text" name="greA" />
            </div>
            <div class="mb-3">
              <label class="input-label">Cumulative GPA of Undergraduate:</label>
              <input class="form-control" type="text" name="cgpa" />
            </div>
            <div class="text-center mt-4">
              <button class="btn btn-primary" type="submit"
value="Submit">Submit</button>
            </div>
          </form>
        </div>
      </div>
      <footer class="footer">
      </footer>
    </div>
</body>
</html>
```

graduate.html:

```
<html>

<head>
  <title>My Project</title>

  <link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/css/bootstrap.min.css"
    integrity="sha384-
JcKb8q3iqJ61gNV9KGb8thSsNjpSL0n8PARn9HuZOnIxN0hoP+VmmDGMN5t9UJ0Z
" crossorigin="anonymous" />
  <script src="https://code.jquery.com/jquery-3.5.1.slim.min.js"
    integrity="sha384-
DfXdz2htPH0lsSSs5nCTpuj/zy4C+OGpamoFVy38MVBnE+IbbVYUew+OrCXaRkfj"
    crossorigin="anonymous"></script>
  <script src="https://cdn.jsdelivr.net/npm/popper.js@1.16.1/dist/umd/popper.min.js"
    integrity="sha384-
9/reFTGAW83EW2RDu2S0VKaIzap3H66lZH81PoYlFhbGU+6BZp6G7niu735Sk7lN"
    crossorigin="anonymous"></script>
  <script src="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/js/bootstrap.min.js"
    integrity="sha384-
B4gt1jrGC7Jh4AgTPSdUtOBvfO8shuf57BaghqFfPlYxofvL8/KUEfYiJOMMV+rV"
    crossorigin="anonymous"></script>
  <style>
    .bgContainer {
      background-color: #f2f4f8;
    }

    .subscribe-form-container
      { background-color: #ffffff;
      border-bottom-left-radius: 8px;
      border-bottom-right-radius: 8px;
    }
```

```
    .input-label
      { color: #7b8794;
      font-family: "Roboto";
      font-size: 12px;
      font-weight: bold;
    }
  </style>

</head>

<body>

  <div class="container">
    <nav class="navbar navbar-expand-md navbar-dark bg-dark">
      <h3 class="navbar-brand"></h3>
      <button class="navbar-toggler" type="button" data-toggle="collapse" data-
target="#navbarsExample05"
        aria-controls="navbarsExample05" aria-expanded="false" aria-label="Toggle
navigation">
        <span class="navbar-toggler-icon"></span>
      </button>
      <div class="collapse navbar-collapse" id="navbarsExample05">
        <ul class="navbar-nav mr-auto">
          <li class="nav-item active">
            <a class="nav-link" href="/main">Home</a>
          </li>
          <li class="nav-item">
            <a class="nav-link" href="/undergraduate">Undergraduate College</a>
          </li>
          <li class="nav-item">
            <a class="nav-link" href="/graduate">Graduate College<span class="sr-
only">(current)</span></a>
          </li>
```

```
      </ul>
    </div>
  </nav>
</div>

<div class="container">
  <div class="jumbotron">
    <h4 style="color: #7b8794;">Graduate Universities</h4>


    <div class="p-4 bgContainer">

      <div class="Subscribe-form-Container d-flex flex-row justify-content-center p-4">


        <form action="/graduatealgo">
          <h6 style="color: #7b8794;">Please Enter your GRE score and GPA</h6>
          <br>
          <div class="mb-3">
            <label class="input-label">GRE Verbal Score:</label>
            <input class="form-control" type="text" name="greV" />
          </div>
          <div class="mb-3">
            <label class="input-label">GRE Quantitative Score:</label>
            <input class="form-control" type="text" name="greQ" />
          </div>
          <div class="mb-3">
            <label class="input-label">GRE Writing Score:</label>
            <input class="form-control" type="text" name="greA" />
          </div>
          <div class="mb-3">
            <label class="input-label">Cumulative GPA of Undergraduate:</label>
            <input class="form-control" type="text" name="cgpa" />
          </div>
```

```
            <div class="text-center mt-4">
              <button class="btn btn-primary" type="submit"
value="Submit">Submit</button>
            </div>

          </form>


        </div>
      </div>

      <footer class="footer">
      </footer>
    </div>
</body>

</html>
```

Undergraduate.html:

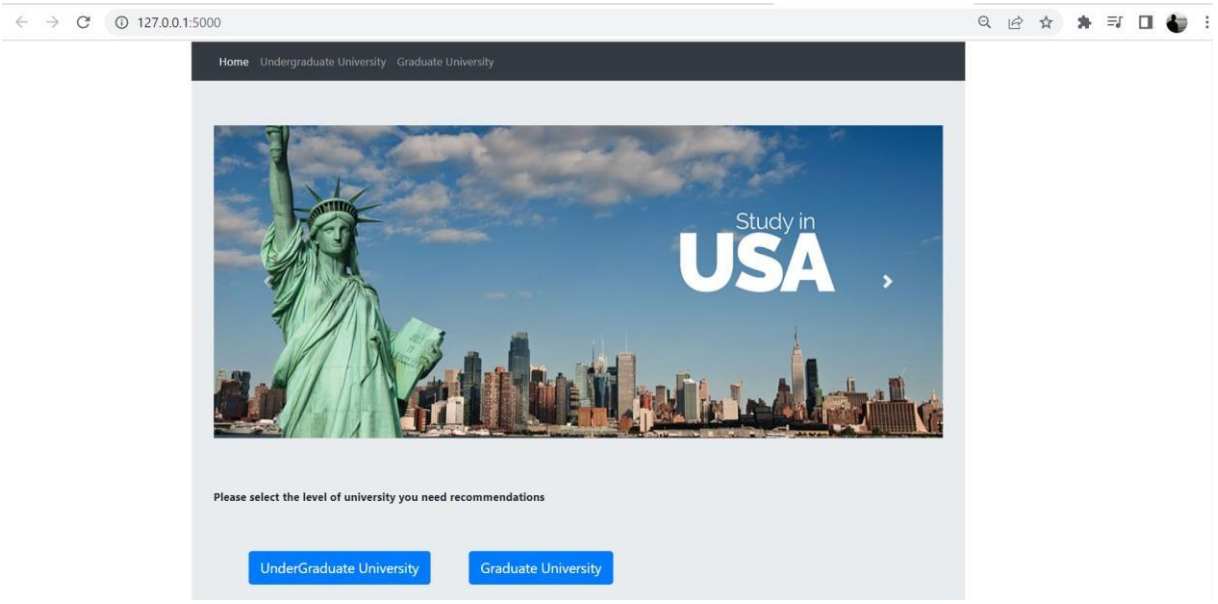
```
<html>
  <head>
    <title>University Recommendation Application</title>
    <link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/css/bootstrap.min.css"
    integrity="sha384-
JcKb8q3iqJ61gNV9KGb8thSsNjpSL0n8PARn9HuZOnIxN0hoP+VmmDGMN5t9UJ0Z
" crossorigin="anonymous" />
  <script src="https://code.jquery.com/jquery-3.5.1.slim.min.js"
    integrity="sha384-
DfXdz2htPH0lsSSs5nCTpuj/zy4C+OGpamoFVy38MVBnE+IbbVYUew+OrCXaRkfj"
    crossorigin="anonymous"></script>
  <script src="https://cdn.jsdelivr.net/npm/popper.js@1.16.1/dist/umd/popper.min.js"
    integrity="sha384-
```

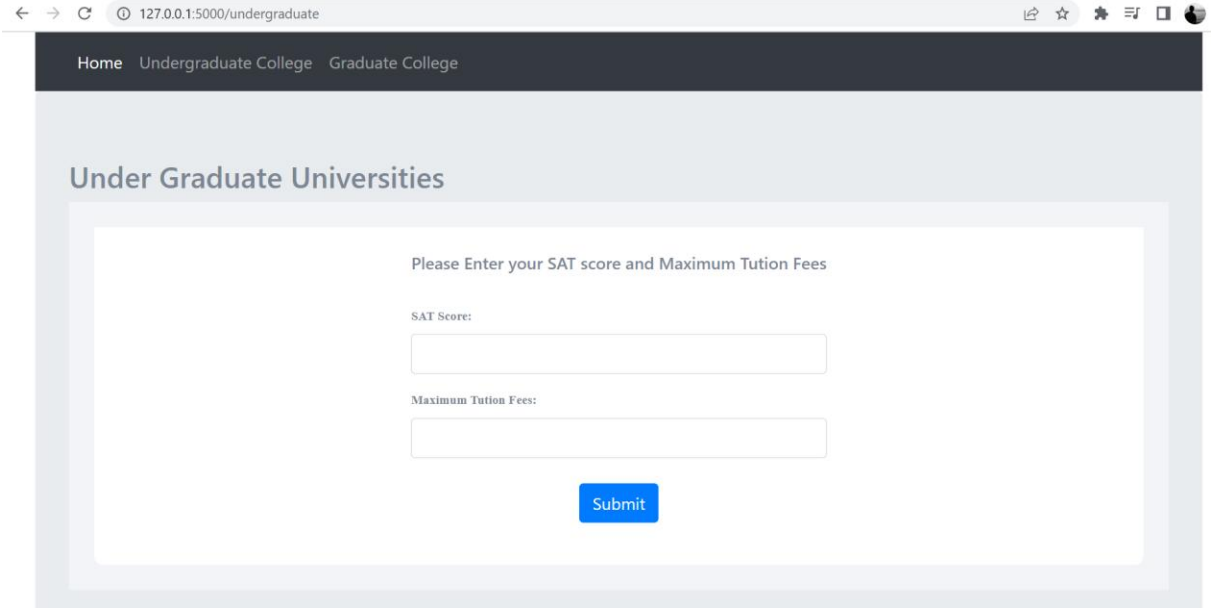
```
9/reFTGAW83EW2RDu2S0VKaIzap3H66lZH81PoYlFhbGU+6BZp6G7niu735Sk7lN"
    crossorigin="anonymous"></script>
  <script src="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/js/bootstrap.min.js"
    integrity="sha384-
B4gt1jrGC7Jh4AgTPSdUtOBvfO8shuf57BaghqFfPlYxofvL8/KUEfYiJOMMV+rV"
    crossorigin="anonymous"></script>
    <style>
      .bgContainer {
        background-color: #f2f4f8;
      }

      .subscribe-form-container
        { background-color: #ffffff;
        border-bottom-left-radius: 8px;
        border-bottom-right-radius: 8px;
      }

      .input-label
        { color: #7b8794;
        font-family: "Roboto";
        font-size: 12px;
        font-weight: bold;
      }
    </style>

</head>
  <body>

    <div class="container">
      <nav class="navbar navbar-expand-md navbar-dark bg-dark">
          <h3 class="navbar-brand"></h3>
          <button class="navbar-toggler" type="button" data-toggle="collapse" data-
target="#navbarsExample05" aria-controls="navbarsExample05" aria-
expanded="false" aria-label="Toggle navigation">
```

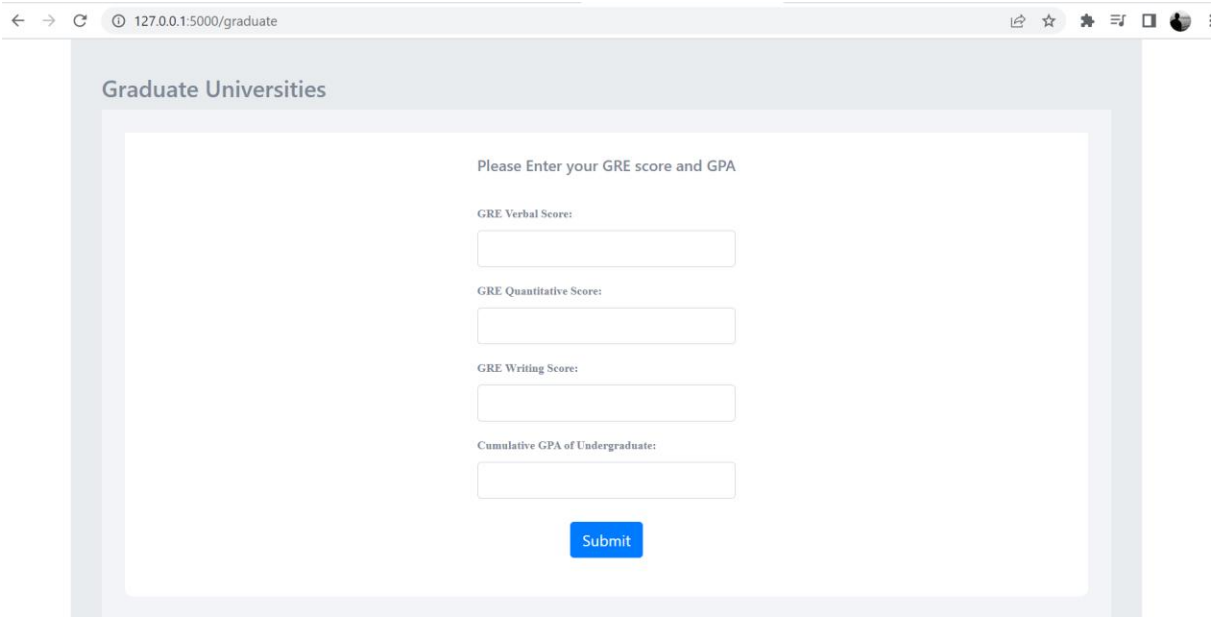
```
            <span class="navbar-toggler-icon"></span>
          </button>
          <div class="collapse navbar-collapse" id="navbarsExample05">
            <ul class="navbar-nav mr-auto">
              <li class="nav-item active">
                <a class="nav-link" href="/main">Home</a>
              </li>
              <li class="nav-item">
                <a class="nav-link" href="/undergraduate">Undergraduate College<span
class="sr-only">(current)</span></a>
              </li>
              <li class="nav-item">
                <a class="nav-link" href="/graduate">Graduate College</a>
              </li>
            </ul>
          </div>
        </nav>
      </div>

      <div class="container">
        <div class="jumbotron">
          <h3 style="color: #7b8794;">Under Graduate Universities</h3>
            <div class="p-4 bgContainer">

              <div class="Subscribe-form-Container d-flex flex-row justify-content-center p-
4">


                <form action="/undergraduatealgo">
                  <h6 style="color: #7b8794;">Please Enter your SAT score and Maximum
Tution Fees</h6>
                  <br>
                  <div class="mb-3">
                    <label class="input-label">SAT Score:</label>
```

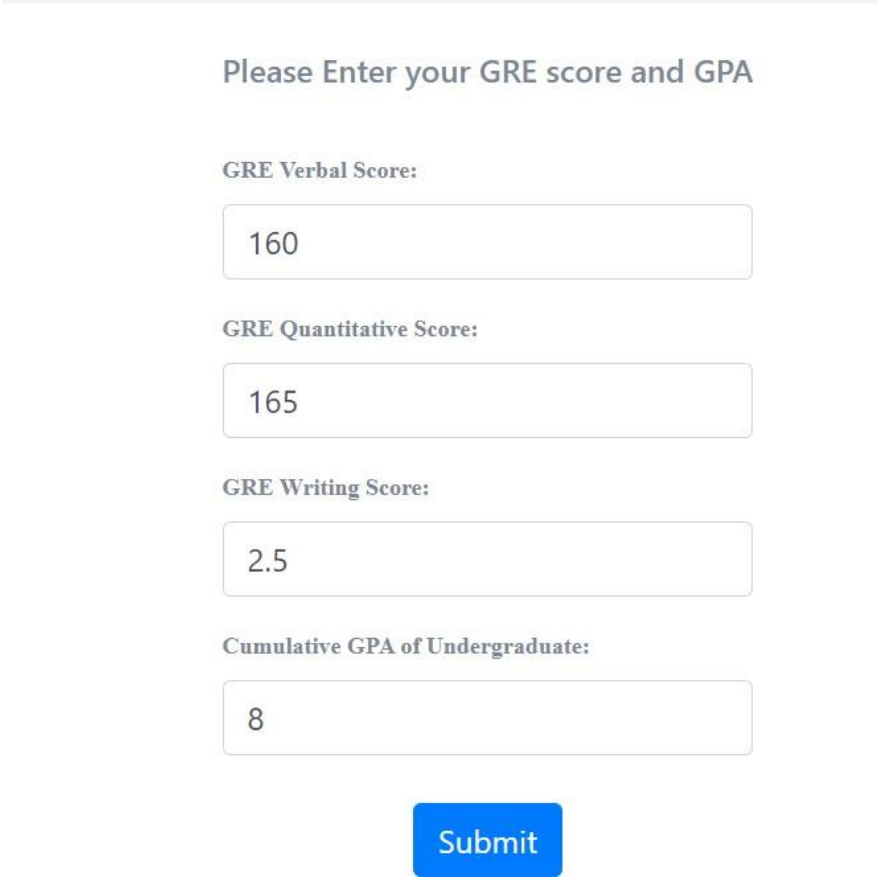
```
                <input class="form-control" type="text" name="sat" required />
              </div>
              <div class="mb-3">
                <label class="input-label">Maximum Tution Fees:</label>
                <input class="form-control" type="text" name="tution" />
              </div>
              <div class="text-center mt-4">
                <button class="btn btn-primary" type="submit"
value="Submit">Submit</button>
              </div>

            </form>


          </div>
        </div>
    </div>

    <footer class="footer">
    </footer>
  </div>
  </body>
</html>
```

B. SCREENSHOTS:






127.0.0.1:5000/graduate
Graduate Universities
Please Enter your GRE score and GPA
GRE Verbal Score:
GRE Quantitative Score:
GRE Writing Score:
Cumulative GPA of Undergraduate:
Submit


Please Enter your GRE score and GPA
GRE Verbal Score:
160
GRE Quantitative Score:
165
GRE Writing Score:
2.5
Cumulative GPA of Undergraduate:
8
Submit

Higher Education Recommendation system

The top recommended Universities based on GRE score and GPA are

S.No	University
1	University Of Southern California
2	Harvard University
3	University Of Pennsylvania (UPenn)
4	University Of Maryland College Park
5	New York University

Under Graduate Universities

Please Enter your SAT score and Maximum Tution Fees

SAT Score:

1500

Maximum Tution Fees:

150000

Submit

Higher Education Recommendation System

The top recommended Universities based on your SAT Score & Maximum Tution Fee are

S.No	University	Acceptance Rate
1	The University of Texas at El Paso	0.9108939326420834
2	Virginia State University	0.8846559533768179
3	Norfolk State University	0.8770874166959707
4	Brigham Young University-Idaho	0.8725509648288466
5	Indiana State University	0.8654836255109772

C. RESEARCH PAPER

HIGHER EDUCATION RECOMMENDATION SYSTEM USING MACHINE LEARNING

Barri Dinesh
Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology Chennai, India
barridinesh10@gmail.com

Kalaga Lokanadham
Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology Chennai, India
Kalagalokanadham143@gmail.com

Jemshia Miriam
Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology Chennai, India
jemshia.cse@sathyabama.ac.in

Abstract - In many groups, machine getting to know techniques are used to analyze large amounts of available facts, providing records for selection-making methods. In the sphere of training, machine gaining knowledge of is used for an expansion of programs including guidelines for students in tenth grade and utility. One of the largest milestones in a person's life is introspection, critical thinking and subsequently making a decision. Using device getting to know algorithms like KNN, XG boost that have expected publications/institutions in real case examine selection aid. Career picks are motivated by using the reviews of mother and father, pals, family, instructors and the media. Hopefully this may help you recognize what each profession path includes and help you analyze and pick out the assets you need to be satisfied and content in your preferred job Today, with more alternatives and ever-growing opposition you need to reflect on consideration on your life accurately and as quickly as viable. While selecting a flow after tenth, schooling path or course and institution, you want to recognize your competencies, and character. In addition, you want to accumulate records approximately distinct profession alternatives, eligibility standards, establishments/colleges and other selection criteria and enterprise markets. The cutting-edge device carried out by way of the Department of Education for diverse instructional and vocational guides is available, as well as instructional institutions through competitive examinations. Based on GRE and SAT scores, this project aims to predict higher education opportunities for students. Using machine learning algorithms, the proposed recommendation system offers personalized recommendations, thereby facilitating the decision-making process and improving the chances of success for students.

Keywords: KNN, XG Boost…

I. INTRODUCTION

There are imperatives for better training to improve the use of statistics. The first is because of outside motives, and the second is due to the continuous improvement of internal nice. The sharp decline in monetary and public assist has precipitated governments to acquire information to help the belief that accountable institutions receive sales. By running defensively, more colleges and universities ought to save you unwanted changes from enjoyable greater records requests. However, at a higher stage, organizations that deliberately use information to improve overall performance are problem to compliance-primarily based necessities, whilst building a statistics-driven destiny. The assumption that get right of entry to statistics has changed in better schooling can be little challenged. At the equal time, it is also clear that era can be used for brand spanking new conversations. New strategies, consisting of analytics or predictive analytics, offer establishments with new ways to use facts to increase their efficiency and better serve students (see, for instance, Bichsel, 2012 and WCET, n.D.). Colleges and universities are coming into generation in which strategic facts about pupil gaining knowledge of and fulfillment, budgeting and performance can be mixed underneath the umbrella of large records. Currently, better education collects greater facts than ever earlier than. But these efforts are usually focused on the first vital, reporting compliance, in place of the second vital, improving institutional design. Advanced organizations will quick remedy this seeming dichotomy. They look for possibilities to construct capability, remove bottlenecks to overcome existing consumer limitations, and locate methods to deliver records and techniques collectively. As a end result, the challenge of the university may be promoted, meeting the requirements of external rules and improving pupil fulfillment. The strategic ideas and statistics that serve the ones techniques come free. In this bankruptcy, we study each the opportunities and obstacles to making

and the use of strategic and operational facts. We also use successful benchmarks based on our revel in with higher training establishments to facilitate strategic planning and create a way of life of studies and documentation. We also discover new technology and their promise of assisting establishments assist college students. This chapter is supposed to provide practical recommendation, no longer a theoretical overview of the concepts of strategic making plans. We need businesses which are strong sufficient to obtain travel facts. To this cease, this bankruptcy also affords recommendation, based totally on private experience and new tendencies in knowledge management, so that you can navigate new paths.

II. OBJECTIVE

One of the main dreams of the newly advanced higher training system is to create a more diverse better education machine with flexible and adaptive institutions. In the sphere of schooling, machine studying is used for a diffusion of programs including recommendations for college students in tenth grade and software. Hopefully this may help you recognize what each profession path includes and help you analyze and pick out the assets you need to be successful, satisfied and content in your preferred profession/profession. While choosing a movement after tenth, education course or path and twelfth organization, you need to recognize your skills, pursuits and character. In addition, you need to accumulate data about extraordinary career alternatives, eligibility criteria, institutions/schools and other choice standards and enterprise markets.

III. LITERATURE REVIEW

Literature assessment is the maximum vital step inside the software development method. Before the tool is evolved, the time thing, the financial system and the electricity of the corporation must be determined. When a majority of these situations are met, the following step is to decide which operating gadget and language may be used to expand the device. When programmers start building a tool, they need numerous external supports. This aid may be acquired from older software, from books, or from websites. Before creating a gadget, those considerations are taken into consideration while the gadget is being advanced. The most a part of the challenge improvement is considering and absolutely studying all the necessities important for the development of the mission. For any cause, literature evaluate is the most essential a part of the software program improvement technique. Before the gear are developed and their associated design, time issue, useful resource necessities, manpower, financial and organisation strengths are recognized and analyzed. With these items happy and fully understood, the next step is to determine the specification of the software in the respective machine, as to what kind of working gadget is required for the cause, and what is needed to move in all the essential software. To the next steps to develop related gear and activities. **[1]** Barriers and drivers of innovation in better training: proof-based totally case studies from ten European universities the article provides to the cutting-edge information about the factors associated with the better schooling organization within the research of innovation in training. Based on the evaluation of ten institutional examples from five European nations, a complete description and class of the bounds and drivers of innovation is given. The results indicate a positive "disunity" within the courting between universities and educational advisers, businessmen and students, in addition to between college leaders and their subjects. Based on the results obtained, the primary issues associated with innovation in better schooling are mentioned, and applicable realistic hints are presented. **[2]** Academic Decision Support in Higher Education Using Machine Learning Algorithms Decisions made by using deans and directors of college research have a terrific impact on the complete instructional community in addition to on society as a whole. In this newsletter, we gift the effects of our findings approximately what instructional decisions rely and what variables are concerned in them. Using device learning algorithms, we anticipated the wide variety of graduates in a actual case take a look at to aid selection making. Real records from five undergraduate engineering programs at the Universidad Francisco José de Caldas in Colombia illustrate our results. A evaluation between vector system and synthetic neural network is made the usage of confusion matrix and receiver performance curve. Methods and algorithm structure are offered. **[3]** The Effects of Blended Learning on Student Performance in a Higher-Level Education Course: A Meta Analysis This article

explains the impact of mixed mastering (BL) on the educational performance of university college students. A meta-evaluation (okay = 51 impact sizes) become performed to statistically synthesize studies that compared student consequences in a BL placing with traditional faculty settings. We include training and teacher evaluation techniques on the end of the path as manage variables. The consequences display that BL has a small internet impact ($g^+ = 0.385$, $p < 0.001$) compared to standard coaching techniques. STEM disciplines were observed to have a significantly better mean impact length ($g^+ =$ zero.496) as compared to non-STEM disciplines. ($r^+ = 0.210$). However, the weighted common effect sizes do now not display big differences between the evaluation strategies, namely unmarried-factor and multi-factor assessment. The locating confirms that BL is significantly associated with higher academic achievement in STEM students than with traditional college exercise. Accordingly, the discussion of the final results and implications for destiny studies is more suitable. **[4]** Using structures dynamics to broaden training for sustainable development in higher schooling with a focal point on sustainable studying capabilities In reaction to the growing public problem approximately the demanding situations of sustainable development and the strengthening of international calls to transport towards a sustainable destiny, better schooling have to be worried inside the implementation of education for sustainable improvement; because of this, college graduates are part of this view as future leaders and heirs of technology. In this examine, the principal research query is: what mechanisms are vital to sell sustainable education in higher education with a focal point on sustainable getting to know talents? It is a dynamic machine studies technique. Therefore, we used a combined studies method. Research tools protected literature evaluation, questionnaires, interviews and statement. We have advanced a dynamic model for educational improvement for sustainable development in higher education with a focus on student competence in sustainable improvement. This version describes the studies problem and predicts the conduct of the model variables thru simulation over the following two decades. This version includes 18 amplifying and six remarks equalizing loops. After the model is cured, the machines are extracted from it. Finally, we evaluated these mechanisms to determine the impact of problem fixing. **[5]** Strategic Day and Night Between Resilience and Information Systems: A Situation Analysis in Public Institutions of Higher Education in Malaysia Higher training establishments have integrated sustainability initiatives into their center enterprise thru curriculum, research, network and operations in reaction to the global shift toward a sustainable destiny. Most research have been performed regarding the overall sustainability of the missions and regulations of better schooling establishments. However, there isn't always enough paintings that presents the primary function of facts systems in helping higher training practices. This truth prompts the study to focus on the vital function of facts systems in enforcing sustainable development. A initial analysis of the state of affairs is completed to see how the institutions of higher schooling implement sustainability initiatives and to affirm expertise gaps in practice. The evaluation become completed at some point of the visits through semi-established interviews with several humans from Malaysian educational institutions that are implementing inside the discipline of sustainability. According to a initial case observe, better schooling establishments do not keep in mind the statistics gadget to support their implementation at the start. No application facts machine to assist their sustainable improvement practices. They still use guide strategies to collect sustainability signs and to evaluate their sustainability performance. As a result, the sustainability of the deliberative development is remoted and it isn't always viable to assess their sustainability results. There are issues related to statistics control and sustainable development approaches. Thus, this has a look at highlights the concept of a strategic dating between sustainability and data structures. Implementing this idea via this examine will amplify our understanding of the important thing position of information structures in supporting sustainable improvement practices and in achieving successful campus sustainability.

IV. EXISTING SYSTEM

The modern device makes use of 3 superior classification algorithms to expect the range of graduates based totally on real statistics on undergraduate college students in South America. Receiver performance and accuracy curve analysis is achieved as a performance degree to compare and

examine selection tree, logistic regression, and random area wherein the latter plays best.

V. PROPOSED SYSTEM

The machine makes use of gadget getting to know for a ramification of programs to offer college students while graduation and gain. One of the biggest milestones in a person's lifestyles is introspection, essential questioning and finally you decide. While choosing a move after tenth, education course or path and twelfth group, you want to recognize your abilities, hobbies and persona. In addition, you must acquire facts approximately distinct profession alternatives, eligibility standards, institutes/schools and different choice standards and enterprise markets.

ADVANTAGES OF PROPOSED SYSTEM

- Accuracy is high.
- High efficiency

MODULES

1) Data Collection Module:

The data is a worldwide facts series. This device uses the Pima Indian dataset to train the model. The records set is the primary set of records that the program normally includes. This is the one wherein we must installation the version first because we must set up the feature and this statistic is available inside the device. This statistic is used to instruct the system in diverse activities. This is the statistics on which a set of rules may be skilled to educate the model and do the paintings automatically. Test statistics is input to the program. It suggests how the information influences the execution of the module it represents, and this is particularly used for checking out.

2) Pre-Processing Module:

Data preprocessing is a procedure that truly uses the underlying transformation information right into a pure dataset. This is the step in which the records is converted or encoded so that the machine can without problems parse it. The important characteristic of statistics preprocessing in the education process is to take away pointless statistics and fill within the missing value. So, the device can be effortlessly mounted to assist.

3) Feature Extraction

Feature extraction is a method in which a secret's used to adjust output traits. This square property is used to calculate the homes of samples that range inside the type of the primary components of the pattern. This technique includes reducing the resources required to explain massive facts set. Feature extraction is the technique of lowering attributes. It also tends to boom the speed and efficiency of the officer's research.

4) Apply Machine Learning Algorithm

The ML algorithm is a non-parametric method proposed by means of Thomas Cover and used for regression and class. This set of rules is particularly used to discover strength issues. A system mastering algorithm is a class-based totally studying method. This set of rules uses a distance to suggest gadgets, via normalizing the training facts to greatly improve its accuracy. K Nearest Neighbors are obtained from a hard and fast of objects for which training or values of gadgets are recognized. It may be concept of as a training set of rules, although no specific education steps are required.

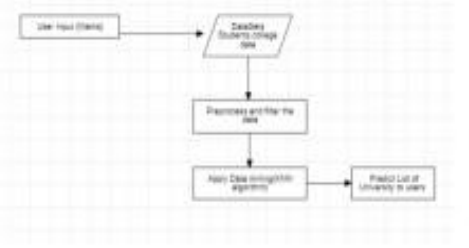
5) Prediction Module

Its module makes use of gadget getting to know for various packages to offer recommendations to students based on GRE, Academic Grades. One of the biggest milestones in a person's existence is introspection, essential questioning and finally you make a decision. While selecting a circulate after tenth, training route or course and twelfth organization, you need to understand your talents, pursuits and character. In addition, you want to acquire information about specific career alternatives, eligibility criteria, institutions/colleges and other choice criteria and business markets.

VI. SYSTEM ARCHITECTURE

A description of the overall characteristics of the program is mixed with a definition of the requirements and a assertion of the higher order. In the architectural layout, the numerous pages and their relationships are recognized and designed. Major software components are identified and damaged down into processing methods and conceptual facts systems, and

relationships among modules are identified. The proposed device includes these modules.



REQUIREMENT ANALYSIS

Requirements evaluation, additionally called necessities development, is the technique of determining person expectations for a new changed product. Includes obligations that determine the need to analyze, file, verify, and manage software systems or requirements. Documented requirements, actionable, measurable, verifiable, and falsifiable, talk over with recognized desires or possibilities, and are defined at a level of detail enough for the motive.

Usability

It suggests how easy the device is. It is straightforward to invite requests in any form, both brief and long, Porter's algorithm stimulates the favored response of the person.

Security

State cozy get entry to to grant aid is protection. A true protection device and unauthorized users cannot get admission to the gadget, they location excessive safety within the secure.

Reliability

This is the opportunity that on every occasion the programmer crashes. The measure is regularly expressed in terms of MTBF (time between failures). It is a essential requirement for the procedure to run efficiently and completely without interruption. It can take care of any load, continue to exist and continue to exist, and even overcome any failure.

Safety

Security is a degree to save you evil. Each request is treated in a at ease manner to prevent others from knowing your private statistics.

NON- FUNCTIONAL REQUIREMENTS

Portability

It is handy to apply the equal software program for diverse purposes. The undertaking can run on any running machine.

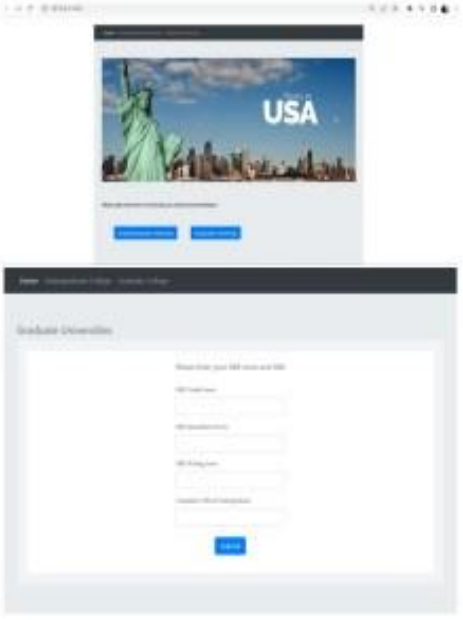
Performance

These necessities define the specified assets, time, space, bandwidth, and the whole thing else associated with gadget performance.

Accuracy

The results of the hunt queries are very correct and *high speed of acquiring facts. The degree of protection provided by the machine is excessive and powerful.

VII. RESULT



VIII. CONCLUSION

In this article, a class hassle in a scholar database is used to predict the student's choice based totally on the preceding database. Since there are several approaches which might be used to categories the facts, k-method clustering classifier and KNN classifier are used right here. Information which includes grades, the use of records from preceding students is gathered to predict the course of look at or profession and twelfth grade at the end. This takes a look at will assist the students to improve the students' selection. This can assist college students enhance their educational understanding, so that it will in the long run cause the proper decision in their careers.

REFERENCES

[1] Lasakova, L.Baizikova and LDedze, "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. Nieto, V. Garcia-Diaz, C. Montenegro, and R.G. Crespo, "Supporting academic decision making at higher educational institutions using machine learning based algorithms," SoftComput.,Vol.23,no.12,pp.4145-4153,2018.

[3] H.M. Vo, C. Zhu, and A. Dieo, "The effect of blended learning on student performance at course level in higher education: Ameta-analysis," stud. Educ. Ecal., vol. 53, pp. 17-28, Jun.2017

[4] E.Faham,A.Rezvanfar,S.H.M.Mohammadi,andM .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.

[5] 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.

[6] Y. V. N. Acevedo and C. E. M. Marín, ''System architecture based on learning analytics to educational decision makers toolkit,'' Adv. Computer. Sci. Eng., vol. 13, no. 2, pp. 89–105, 2014.

[7] R. T. Góes, M. T. A. Steiner, and P. J. S. Neto,Educationqualitymeasuredbytheclassificatio nofschoolperformanceusingqualitylabes,'' Appl. Mech. Mater., vols. 670–671, pp. 1675–1683, Oct. 2014.

[8] H.LounisandT.Fares,''Usingefficientmachine-learningmodelstoassess 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.

[9] K. V. Pincus, D. E. Stout, J. E. Sorensen, K. D. Stocks, and R. A. Lawson, ''Forcesforchangeinhighereducationandimplicati onsfortheaccounting academy,'' J. Accounting Educ., vol. 40, pp. 1–18, Sep. 2016.

[10] 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.

[11] 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.

[12] M. Taucean and M. Tamasila, ''Research challenges for eLearning supportinengineeringandmanagementtraining,''P rocedia-SocialBehav. Sci., vol. 124, pp. 210–218, Mar. 2014.

[13] I. González-González and A. I. Jiménez-Zarco, ''Using learning methodologies and resources in the development of critical thinking competency: