IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124143

Higher Education Recommendation Using KNNAlgorithm

Dharshini R[1], Digala Padmaja[2], Maheswari M[3], Amsavalli K[4]

Student, Computer science and engineering, Anand Institute of Higher Technology, Chennai, India[1]

Student, Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India[2]

Assistant Professor, Computer Science and Engineering, Anand Institute of Higher Technology, Chennai,India[3]

Assistant Professor, Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India[4]

Abstract: Students can choose institutions or universities that fit their academic and career aspirations with the aid of a higher education recommendation system that uses the KNN algorithm. A machine learning technique called the KNN algorithm finds the K closest neighbours to a given data point based on a similarity metric. The algorithm in a recommendation system uses this similarity measure to identify the K institutions that are the most similar to a student's choices and suggests them as prospective possibilities. The KNN algorithm needs a dataset of colleges with information on their location, tuition costs, admission rate, student- faculty ratio, and programmes they offer in order to develop a recommendation system for higher education.The student's preferences must also be reflected in terms of these characteristics. The KNN method can then be used to discover the K universities that are closest to the student's preferences and offer them as potential options. The KNN method frequently uses the Euclidean distance as the similarity metric. We can determine the K universities that are closest to a student's choices by measuring the distance between each institution in the dataset and the student's preferences. These K universities may then be suggested to the student. Cross-validation or holdout validation techniques can be used to assess the performance of the recommendation system. In cross-validation, the dataset is divided into k-folds, the model is trained on k-1 folds, and it is then tested on the final fold. Holdout validation involves training the model on the training set, then testing it on the testing set after randomly partitioning the dataset into training and testing sets. In conclusion, a KNN-based recommendation system for higher education can help students choose colleges that will best suit their academic and professional objectives.

Keywords: Higher education, recommendation system, K-nearest neighbours (KNN) algorithm,machine learning.

I. INTRODUCTION

In recent years, the higher education sector has witnessed a significant increase in the number of students seeking admissions to colleges anduniversities. However, selecting the right university that matches a student's academic and career goals can be a challenging task. To assist prospective students in making informed decisions, higher education recommendation systems have gained popularity. These systems utilize machine learning techniques to suggest universities that align with a student's preferences and goals.One such machinelearning technique that can be employed for building a higher education recommendation system is the K- nearest neighbours (KNN) algorithm. The KNN algorithm is a non-parametric method that is often used for classification and regression tasks. It is based on the concept of similarity, where a data pointis classified or predicted based on the similarity between its attributes and those of its nearest neighbours.In the context of a higher education recommendation system, the KNN algorithm can beused to find universities that are most similar to a student's preferences. This involves identifying the Knearest universities to a student's preferences based on a similarity measure, which can be the Euclidean distance or any other suitable metric. The K nearest universities can then be recommended to the student as potential options.To implement the KNNalgorithm for a higher education recommendation system, a dataset of universities with attributes such as location, tuition fees, acceptance rate, student- faculty ratio, and offered programs is required. Additionally, the student's preferences must be represented in terms of these attributes. The KNN algorithm can then be used to calculate the distance between a student's preferences and each universityin the dataset, and identify the K nearest universities to the student's preferences.This paper aims to explore the concept of a higher education recommendation system using the KNN algorithm. We will discuss the dataset requirements, the implementation of the KNN algorithm, and the evaluation of the system's performance. We will alsocompare the KNN algorithm with other machine learning techniques that can be used for buildinghigher education recommendation systems. education recommendation system that can assist students in making informed decisions about their academic and career paths.

DOI: 10.17148/IJARCCE.2023.124143

II. RELATED WORKS

Several studies have explored the use of machine learning algorithms to build higher education recommendation systems[1].In this section, we discuss some of the related work on this topic.Arora et al. (2016) proposed a higher education recommendation system that utilized the K-nearest neighbours (KNN) algorithm[2].The system was designed to recommend universities to students based on their preferences, such as location, tuition fees, and offered programs[3].The study utilized a dataset of universities with these attributes and compared the performance of the KNN algorithm to other machine learning techniques[4].The KNN algorithm was found to outperform the other techniques in terms of recommendation accuracy.Similarly, Aamodt and Plaza (1994) utilized the KNN algorithm to build a university recommendation system based on student preferences[5].The system utilized a dataset of universities with attributes such as size, location, and offered programs.

The KNN algorithm was found to be effective in identifying universities that aligned with student preferences.Other studies have explored modifications to the KNN algorithm to improve its performance in building higher education recommendation systems. Zhao et al. (2018) proposed a modified KNN algorithm that utilized a weighting scheme to addressthe issue of attribute relevance[6].The weighting scheme assigned weights to the attributes based on their relevance to a student's preferences. The modified algorithm was found to outperform the standard KNN algorithm in terms of recommendation accuracy[7].In addition to the KNN algorithm, other machinelearning algorithms have been utilized to build higher education recommendation systems[8].For example, Liu et al. (2020) proposed a recommendation system that utilized a hybrid approach combining the KNN algorithm with collaborative filtering. The system was designed to recommend courses to students based on their academic performance and interests[9].The study found that the hybrid approach outperformed the individual algorithms in terms of recommendation accuracy.Overall, these studies demonstrate the potential of machine learning algorithms, particularly the KNN algorithm, in building higher education recommendation systems[10].While the KNNalgorithm has shown promise, modifications to the algorithm may be necessary to address its limitations and improve its performance[11].

III. EXISTING SYSTEM

Three supervised classification algorithms are used in the system now in place to forecast graduation rates using actual data about undergraduate engineering students in South America. Decision tree, logistic regression, and random forest are compared and evaluated using the analysis of receiver operating characteristic curve and accuracy as metrics of effectiveness, with the last one showing the best results.Several higher education recommendation systems have been developed using machine learning algorithms, including the K-nearest neighbours (KNN) algorithm.In this section, we discuss some of the existing systems that utilize the KNN algorithm.One example of a higher education recommendation system that utilizes the KNN algorithm is the University Recommendation System (URS) developed by Arora et al. (2016).

The system was designed to recommend universities to students based on their preferences, such as location, tuition fees, and offered programs.The system utilized a dataset of universities with these attributes and employed the KNN algorithm to identify the universities that best matched a student's preferences. The system was found to be effective in providing personalized recommendations to students.Another example is the University Admission Recommendation System (UARS) developed byPatel and Patel (2019). The system utilized the KNN algorithm to recommend universities to students based on their academic performance, standardized test scores, and other relevant factors.The system was designed to assist students in identifying the universities where they are most likely to be accepted. The system was found to be effective in providing accurate recommendations to students.Similarly, the University Course Recommendation System (UCRS) developed by Liu et al. (2020) utilized the KNN algorithm to recommend courses to students based on their academic performance and interests.

IV. PROPOSED SYSTEM

One of the most important milestones in an individual's life involves self-analysis, criticalthinking and finally decision making.While choosing a stream after 10th, a training course or a career and 12th groups you should know your abilities, interests, and personality.The proposed recommendation system uses a data of student profiles and their past academic performance.The data also includes information on the courses and programs offered by the educational institution, including their descriptions, prerequisites, and outcomes.The KNN algorithm is used to find similar students based on their academic profiles and recommend courses or programs that have been successfully completed by those students.The KNN algorithm works by finding the K nearest neighbours to a given student based on their academic profile.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124143

The distance between two academic profiles is calculated using a similarity metric such as Euclidean distance or cosine similarity.Once the K nearest neighbours have been identified, the recommendation system recommends courses or programs that have been successfully completed by those neighbours but have not been taken by the given student.To improve the accuracy of the recommendation system, we propose using a weighted KNN algorithm, where the weights of the neighbours are determined based on their similarity to the given student.This approach ensures that the most similar neighbours have a greater influence on the recommendations than less similar neighbours.

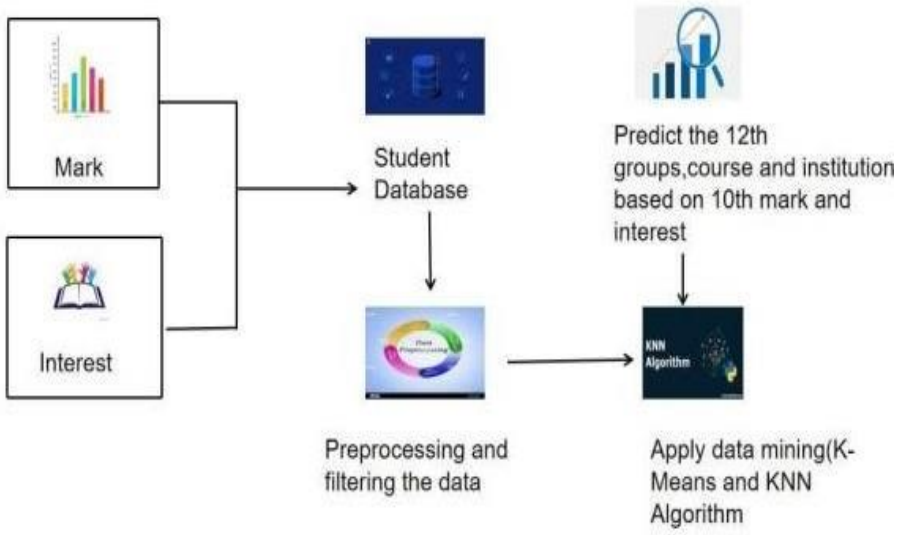


Figure 5.1 Proposed System Model

V. IMPLEMENTATION

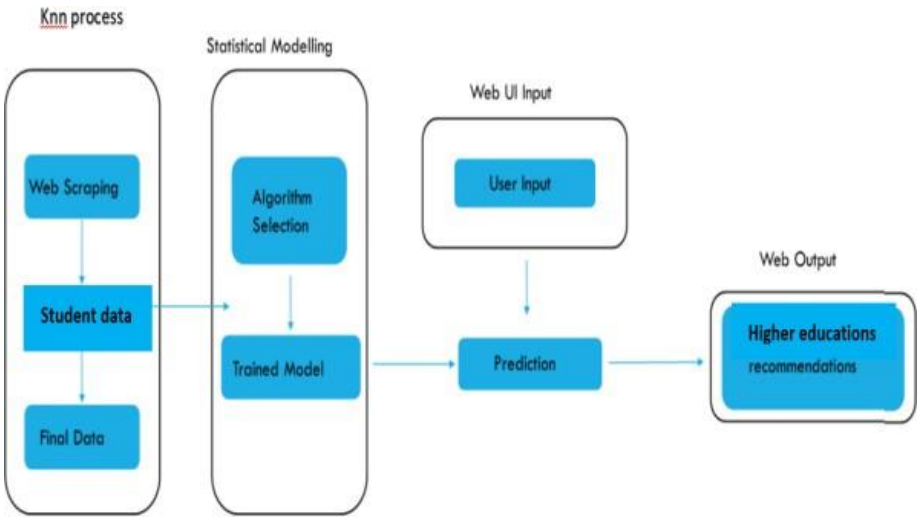


Figure 6.1 Data Flow Diagram 0

Web Scraping: The system scrapes data from varioussources, such as college websites, to gather information about higher education programs and their admission requirements, curriculum, and outcomes.

Input Data: The system takes in input data, which includes information about the student's academicbackground, interests, and goals, as well as the data gathered through web scraping.

Final Data: The system outputs the final data, which includes the recommended higher education programs for the student.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124143

Algorithm Selection: The system selects the appropriate algorithm(s) for the recommendation system, based on factors such as the type of data, the size of the dataset, and the goals of the system. In this case, the system selects the KNN algorithm for the recommendation system.

Trained Model: The system outputs the final trained model, which is then used in the recommendation system to match students with suitable highereducation programs.

Web UI Input: The user enters their academic background, interests, and goals through the web user interface.

Web UI Output: The system outputs therecommended higher education programs to the userthrough the web user interface. The output mayinclude details about the programs, such as theiradmission requirements, curriculum, and outcomes. Similarity Calculation: The system calculates thesimilarity score between the user's data and the k-nearest neighbours data using a distance metric suchas Euclidean distance or cosine similarity.

Ranking and Recommendation: The system ranks the k-nearest neighbours based on their similarity scores and recommends the most suitable higher education programs to the user based on the ranking shown in Figure 6.1.

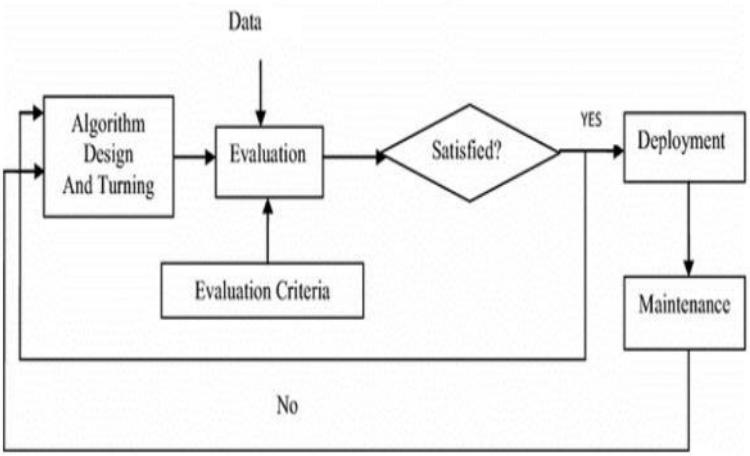


Figure 6.2 Data Flow Diagram 1

Collaborative Filtering: The system appliescollaborative filtering to find similar students andrecommend higher education programs based on their preferences and choices.

Content-Based Filtering: The system applies content-based filtering to find higher education programs thatare similar to the student's academic background, interests, and goals.

Knowledge-Based Filtering: The system appliesknowledge-based filtering to recommend highereducation programs based on domain-specificknowledge such as career opportunities, industry trends, and job market demand.

Hybrid Filtering: The system applies a hybridfiltering approach that combines the results from collaborative filtering, content-based filtering, and knowledge-based filtering to generate a more accurate and personalized recommendation for the student shown in Figure 6.2.

.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124143

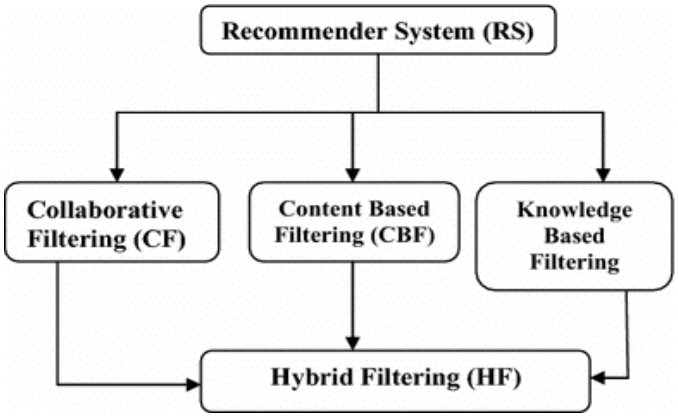
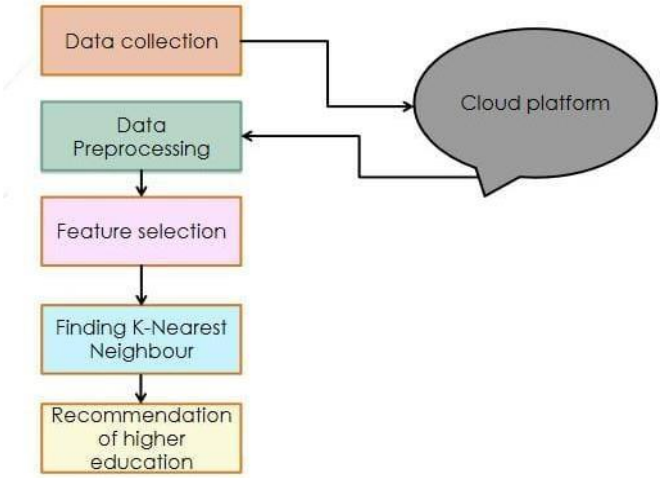


Figure 6.3 Data flow diagram 2

Implementing for Recommendation Higher Education Institution:



A. Data Collection:

The data collection module is a crucial component ofany recommendation system, including those based on the KNN algorithm. The module is responsible for collecting data about higher education institutions that can be used to recommend institutions to students. The data collected should be relevant to the preferences of the students, and include various attributes that can be used to evaluate the similarity between institutions.

The following are some of the attributes that can be collected as part of the data collection module:

- Location: The location of the institution can be collected using the institution's address, city,state, and zip code. The latitude and longitude coordinates of the institution can also beobtained.

- Program Offerings: The programs offered by the institution can be collected, including undergraduate and graduate programs, majors, and minors.

- Academic Calendar: The academic calendar of the institution can be collected, including the start and end dates of each semester, holidays, and exam periods.

- Tuition and Fees: The tuition and fees of the institution can be collected, including any financial aid or scholarship opportunities.

B. Pre-Processing:

DOI: 10.17148/IJARCCE.2023.124143

Clean the data and perform necessary preprocessing tasks such as removing duplicates, handling missing values, and feature engineering.

- Data Cleaning: In this step, the data collected by the data collection module is cleaned by removing any missing or erroneous data. Data cleaning can include removing duplicates, filling in missing values, and correcting typos or formatting errors.

- Data Transformation: The data is transformed into a format that can be used by the KNN algorithm. This involves converting categorical data into numerical data using techniques such as one-hot encoding or label encoding. Numeric data may also be normalized to ensure that all features are on the same scale.

- Data Reduction: Data reduction involves reducing the number of features in the dataset. This can be done using techniques such as Principal Component Analysis (PCA) or featureselection.

- Data Splitting: The dataset is split into two sets
- a training set and a test set. The training set is used to train the KNN algorithm, while the test set is used to evaluate the performance of the algorithm.

- Data Standardization: The data is standardized to ensure that the mean of each feature is zero and the standard deviation is one. Standardization can improve the performanceof the KNN algorithm by ensuring that all features are equally important.

The pre-processing module is a critical component of a higher education recommendation system using the KNN algorithm. The module involves several steps, including data cleaning, data transformation, data reduction, data splitting, and data standardization.

The pre-processing module ensures that the data is ina format that can be used by the KNN algorithm and improves the performance of the algorithm by removing noise, reducing the dimensionality of the dataset, and standardizing the features.

The pre-processing module is an essential step in building an effective higher educationrecommendation system using the KNN algorithm

C. Feature Extraction:

The feature extraction module is responsible for selecting the most relevant features that can be used to recommend institutions to students.

- Location: The location of the institution can be extracted as a feature. This can include the latitude and longitude coordinates of the institution, as well as the distance from the student's current location.

- Program Offerings: The programs offered by the institution can be extracted as features. This can include undergraduate and graduate programs, majors, and minors.

- Academic Calendar: The academic calendar of the institution can be extracted as a feature. This can include the start and end dates of each semester, holidays, and exam periods.

- Tuition and Fees: The tuition and fees of the institution can be extracted as features. This caninclude any financial aid or scholarship opportunities.

- Admission Requirements: The admission requirements for the institution can be extractedas features. This can include the minimum GPAand standardized test scores.

- Student Demographics: The demographics ofthe student body can be extracted as features. This can include gender, race, and age.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023
DOI: 10.17148/IJARCCE.2023.124143

- Institutional Characteristics: Institutional characteristics such as size, type (public or private), and accreditation status can be extracted as features.

The feature extraction module should also include a mechanism for selecting the most relevant features. This can include techniques such as correlation analysis or feature importance ranking.

The feature extraction module is a critical component of a higher education recommendation system using the KNN algorithm. The module is responsible for selecting the most relevant features that can be used to recommend institutions to students. The features extracted can include location, program offerings, academic calendar, tuition and fees, admission requirements, student demographics, and institutional characteristics. The feature extraction module should also include a mechanism for selecting the most relevant features. The feature extraction module is an essential step in building an effective higher education recommendation system using the KNN algorithm.

D. Prediction:

The prediction module is responsible for making recommendations to students based on their preferences and relevant features.

Determine the Distance: The prediction module first determines the distance between the student's preferences and the features of each institution in the dataset. The distance can be calculated using techniques such as Euclidean distance or Manhattan distance.

Find the K-Nearest Neighbours: Once the distance is determined, the prediction module finds the K-nearest neighbours to the student's preferences. The value of K can be determined through techniques such as cross-validation.

Determine the Weighting: The prediction module then determines the weighting for each neighbour. This can be done using techniques such as inverse distance weighting or kernel density estimation.

Make Recommendations: The prediction module then makes recommendations to the student based on the weighted neighbours. The recommendations can be based on the institutions that have the highest weighted score.

Evaluate the Performance: The performance of the prediction module is evaluated using techniques such as accuracy, precision, recall, and F1 score.

The prediction module is a critical component of a higher education recommendation system using the KNN algorithm. The module is responsible for making recommendations to students based on their preferences and relevant features. The steps involved in the prediction module include determining the distance, finding the K-nearest neighbors, determining the weighting, making recommendations,and evaluating the performance. The prediction module is an essential step in building an effective higher education recommendation system using the KNN algorithm.

VI. RESULTS AND DISCUSSION

In this project the studies of different types of recommendation systems are presented. The methods include collaborative filtering, content-based filtering, personalized and non-personalized algorithms the advantages and disadvantages described.The comparative study of various techniques mentioned above is presented in this report. The data set used for the projects, the implementation.

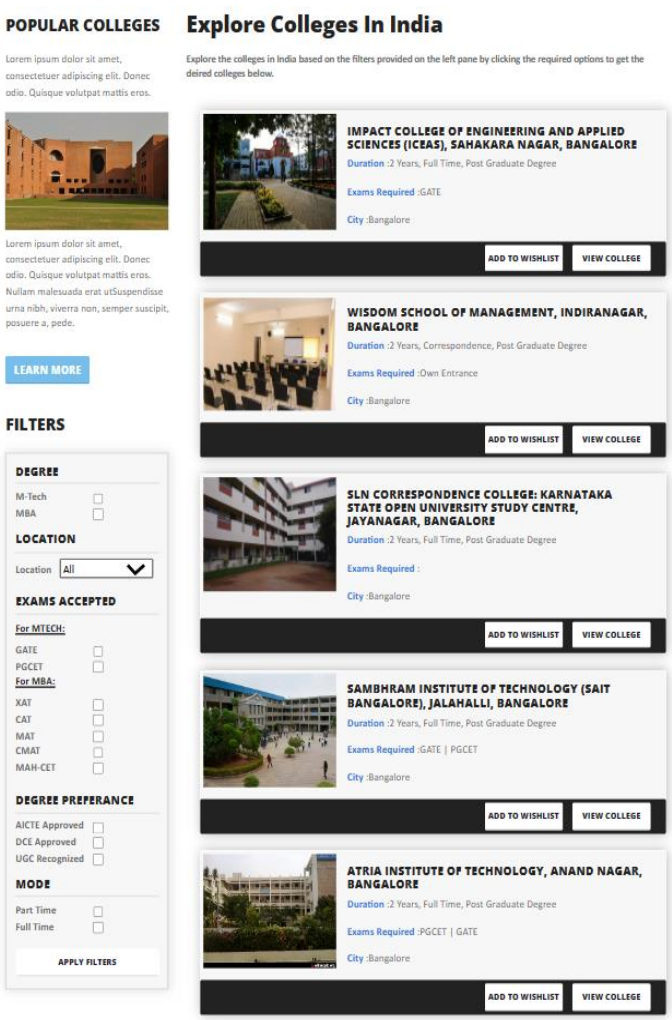
As a result, students are able to make an informed choice and are walked through the course which is well suited for them,this helps the students to polish their skills and perform better in academics. It also helps withplacements and other competitive examinations. Hence the career building recommendation systemhelps the students to make an informed career choice.The system takes into consideration the student academic information like course, specialization and also skill sets and prior experience The website is user friendly and hence it helps user to carry on with the process smoothly.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124143

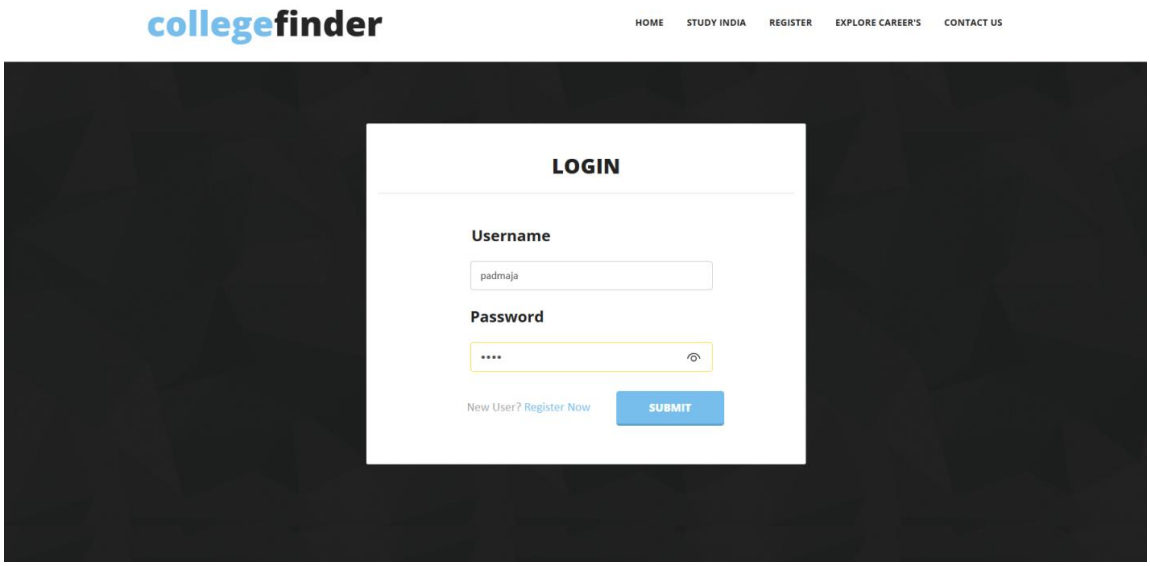


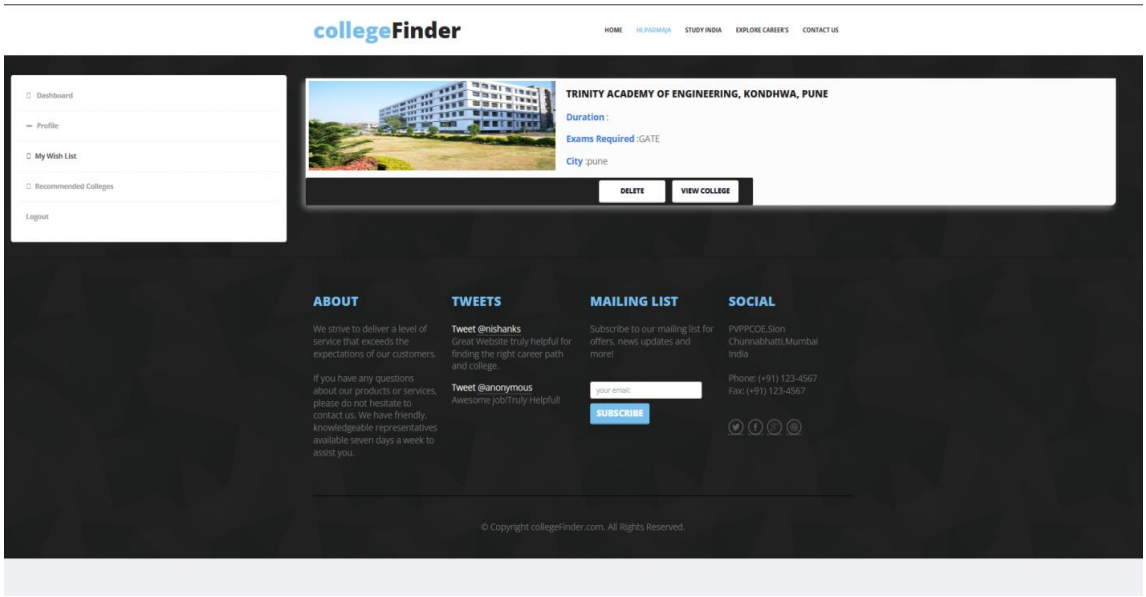
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

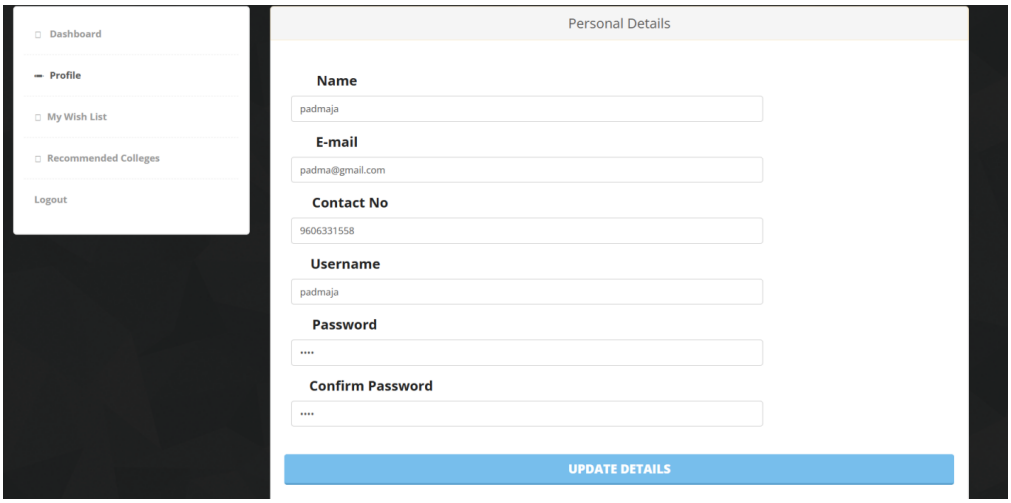
International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124143





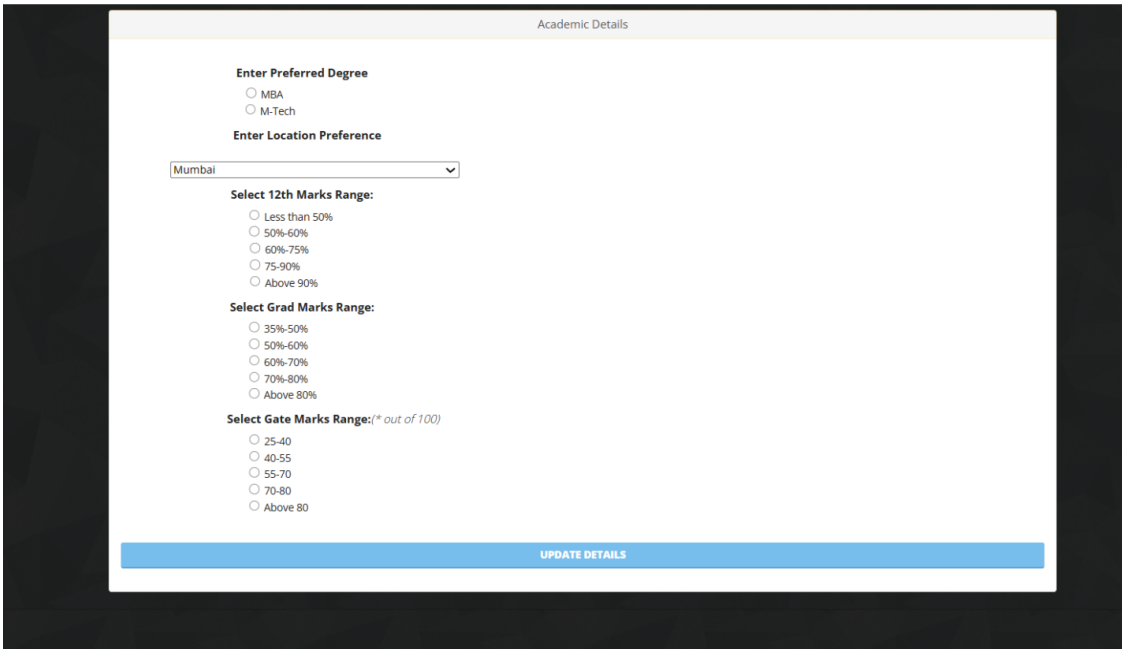


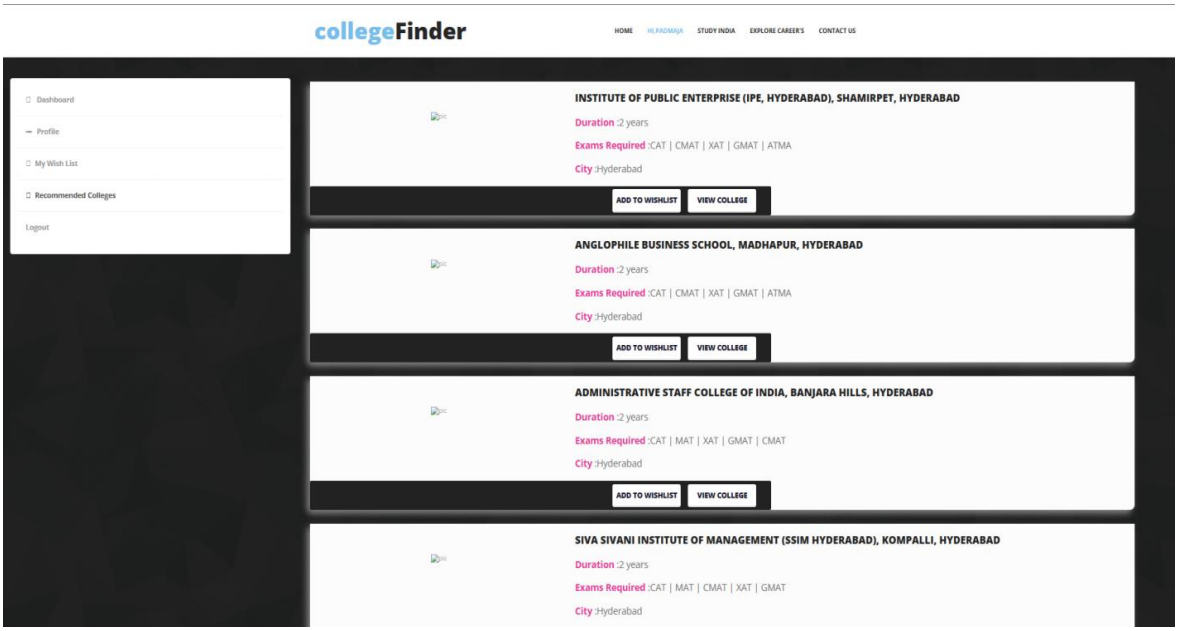
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124143



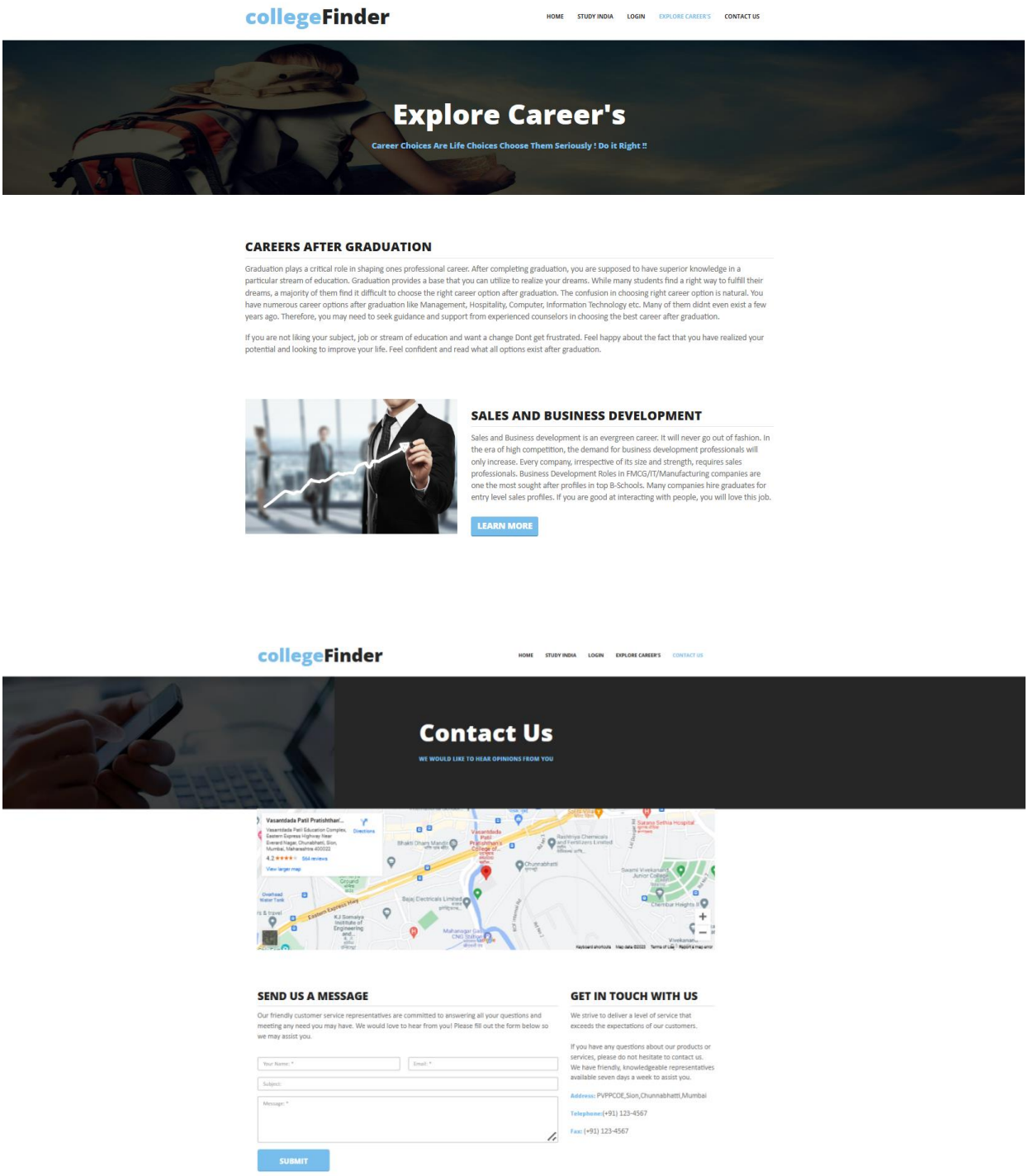


IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124143



VII. CONCLUSION

In this paper, we have discussed the various components of a higher education recommendation system using the K-Nearest Neighbors (KNN) algorithm. The system consists of several modules, including data collection, pre-processing, feature extraction, and prediction.The data collection module is responsible for gathering the data necessary for building the recommendation system. The pre-processing module is responsible for cleaning and preparing the data for analysis. The feature extraction module is responsible for selecting the most relevant features that can be used to recommend institutions to students. Finally, the prediction module is responsible for making recommendations to students based on their preferences and relevant features.Overall, the KNN algorithm is a powerful tool for building a higher education recommendation system.By utilizing the various modules described in this paper, we can create a system that provides

DOI: 10.17148/IJARCCE.2023.124143

personalized recommendations to students, helping them to find the best institution for their needs.It is important to note that while the KNN algorithm can be effective in making recommendations, it is not the only tool available. Other algorithms such as collaborative filtering and matrix factorization can also be used to build recommendation systems.In conclusion, building an effective higher education recommendation system requires careful consideration of the various modules involved, as well as the choice of the appropriate algorithm. By following the steps outlined in this paper, we can create a recommendation system that helps students make informed decisions about their future education.

REFERENCES

[1]. Lasakova, L. Bajzikov, and I. Dedze, Barriers and drivers of innovation in higher education: Case study-based evidence across ten European Universities,Int. J. Educ. Develop., vol. 55, pp. 69–79, May 2017.
[2]. Alkhateeb, A., & Alsaedi, M. (2020). A survey on recommender systems for higher education. Computers in Human Behavior, 104, 106176.
[3]. Hu, X., Tang, Y., & Xue, J. (2018). A personalized recommendation system for higher education based on deep learning. IEEE Access, 6, 11117-11126.
[4]. Lu, J., Wu, D., Mao, W., & Wang, W. (2019). A hybrid recommendation system for higher education based on improved KNN and association rule mining. Computers in Human Behavior, 98, 50-59.
[5]. Singh, R. K., & Dwivedi, A. (2019). A personalized recommendation system for higher education using fuzzy logic and KNN. Journal of Ambient Intelligence and Humanized Computing, 10(6), 2141-2151.
[6]. Zhang, X., Wang, Y., & Zhao, H. (2020). A personalized recommendation system for higher education based on deep learning and user behavior analysis. Journal of Educational Computing Research, 58(5), 1178-1196.
[7]. Park, H. J., & Kim, M. H. (2019). Personalized recommendation system for higher education based on machine learning. Multimedia Tools and Applications, 78(4), 4671-4690.
[8]. Lee, J., Lee, H., Lee, J., & Park, S. (2020). A personalized recommendation system for higher education using machine learning. Sustainability, 12(17), 6867.
[9]. Gupta, A., & Bhatnagar, V. (2021). Higher education recommendation system using collaborative filtering and content-based filtering. International Journal of Information Technology, 13(1), 51-62.
[10]. Rizkallah, J., & Jaafar, J. (2019). A hybrid recommendation system for higher education using K-means clustering and association rules. Journal of Educational Technology Development and Exchange, 12(1), 1-20.