Counsellor – A MERN Stack Recommending System for Educational Institutes

Project Team

Muhammad Abdul Nafay (19P-0117)

Sheheryar Ali (19P-0120)

Saad Javed (19P-0111)

Session 2019-2023

Supervised by

Shoaib Muhammad Khan

Department of Computer Science

National University of Computer and Emerging Sciences

Peshawar, Pakistan

October, 2022

Abstract

One of the biggest challenges for every young adult is to find the right educational institutes according to their preference and interest. Cases of students dropping out at the beginning of the semester are also quite significant. The main cause of failure is the selection of the right institutes that fits the user's preferences. The traditional method of getting recommendation is to either search global universities ranking on the internet or asking other peers for their suggestions. This way is inefficient, unreliable and it also doesn't keep the user's interests in mind. If an overseas student is thinking over further education in Pakistan, then he/she will have a hard time of which university to choose on. This project aims to solve all of these problems by having a platform which requires all the relevant input from the user (such as a user's transcript, interests, budget, location etc.), and it applies different machine learning algorithms to give an output in the form of a list of different institutes that are ranked according to the user's input along with the features of each institute which tells that why it is being recommended.

Table Of Contents

Chapter 1

Introduction

Recommender systems can be defined as programs which attempt to recommend the most suitable items (products or services) to particular users (individuals or businesses) by predicting a user's interest in an item based on related information about the items, the users and the interactions between items and users [1]. The aim of developing recommender systems is to reduce information overload by retrieving the most relevant information and services from a huge amount of data, thereby providing personalized services. The most important feature of a recommender system is its ability to "guess" a user's preferences and interests by analyzing the behavior of this user and/or the behavior of other users to generate personalized recommendations [2].

History of Recommendation System

Recommender Systems (RSs) is an emerging research field that has grown fast and become popular. The increase of interest in this research topic has also been driven by great improvements in internet technology and e-commerce. E-commerce has been one of the major reasons why this became so popular. It helped buyers with no experience in online shopping, by cross-selling the products and by improving customer loyalty. The peak explosion of research in RSs occurred when Amazon launched their Collaborative Filtering (CF) method at the end of the 1990s, successfully increasing their sales. The successful Amazon became popular and other online businesses started to implement RSs on their website. Amazon has patented its CF method as a United States Patent. Due to the fact that the main goal of an RS is to find the preferred information and eliminate information which is not liked by a user, the RS field can be considered as a subset of information filtering. The process of exploring a user's preferences from their historical data is followed by processing it using machine learning algorithms to build a ranked list of recommended items, as preferred by the user [7].

The idea of exploiting computers to recommend the best item for the user has been around since the beginning of computing. The first implementation of the RS concept appeared in 1979, in a system called Grundy, a computer-based librarian that provided suggestions to the user on what books to read. This followed in the early 1990s with the launch of Tapestry, the first commercial RS. Another RS implementation for helping people find their preferred articles was launched in the early 1990s by GroupLens, a research lab at the University of Minnesota, USA. They named the system after the group, GroupLens Recommender System. This system claims to have a similar spirit to that of Tapestry, Ringo, BellCore and Jester. Further development of RSs in the late 1990s was the implementation of Amazon Collaborative Filtering, one of the most widely known RS technologies. Since this era, RSs based on Collaborative Filtering has become very popular and has been implemented by many e-commerce and online systems. Many toolboxes for RSs have also been developed. The success story of Amazon also gave rise to the development of many RS algorithms known as hybrid approaches, which combines collaborative filtering with content-based filtering [7].

Following the successful era at the end of the 1990s, the industry offered generous funding to implement RSs research. The most popular competition in RSs was held by Netflix. They launched the Netflix Prize in 2006 and give a million US Dollars to the winner of the competition who provided the best RS movie recommendation. In 2010, YouTube also implemented an RS on its website and it tries to predict what a user would like to see next based on what they usually like to watch, based on their own preferences and interests [6].

Types of Recommendation System

Content-based recommendation techniques

Content-based (CB) recommendation techniques recommend articles or commodities that are similar to items previously preferred by a specific user [10].

The basic principles of CB recommender systems are:

1) To analyze the description of the items preferred by a particular user to determine the principal common attributes (preferences) that can be used to distinguish these items. These preferences are stored in a user profile.

2) To compare each item's attributes with the user profile so that only items that have a high degree of similarity with the user profile will be recommended [10].

In CB recommender systems, two techniques have been used to generate recommendations. One technique generates recommendations heuristically using traditional information retrieval methods, such as cosine similarity measure. The other technique generates recommendations using statistical learning and machine learning methods, largely building models that are capable of learning users' interests from the historical data (training data) of users.

Collaborative Based Filtering

Collaborative filtering (CF)-based recommendation techniques help people to make choices based on the opinions of other people who share similar interests [11]. The CF technique can be divided into user-based and item-based CF approaches [12]. In the user-based CF approach, a user will receive recommendations of items liked by similar users. In the item-based CF approach, a user will receive recommendations of items that are similar to those they have loved in the past. The similarity between users or items can be calculated by Pearson correlation-based similarity [13], constrained Pearson correlation (CPC)-based similarity, cosine-based similarity, or adjusted cosine-based measures. When calculating the similarity between items using the above measures, only users who have rated both items are considered. This can influence the similarity accuracy when items which have received a very small number of ratings express a high level of similarity with other items. To improve similarity accuracy, an enhanced item-based CF approach was presented by combining the adjusted cosine approach with Jaccard metric as a weighting scheme. To compute the similarity between users, the Jaccard metric was used as a weighting scheme with the CPC to obtain a weighted CPC measure [14]. To deal with the disadvantage of the single-rating based approach, multi-criteria collaborative filtering was developed [15].

Hybrid recommendation

To achieve higher performance and overcome the drawbacks of traditional recommendation techniques, a hybrid recommendation technique that combines the best features of two or more recommendation techniques into one hybrid technique has been proposed. According to Burke [16], there are seven basic hybridization mechanisms of combinations used in recommender systems to build hybrids: weighted, mixed, switching, feature combination, feature augmentation, cascade and meta-level. The most common practice in the existing hybrid recommendation techniques is to combine the CF recommendation techniques with the other recommendation techniques in an attempt to avoid cold-start, sparseness and/or scalability problems.

Motivation

The motivation behind this project is to help young adults decide which educational institute is best according to their interest, academic profile, geo-graphical location and most importantly budget. This is a crucial decision in a student's life, since a wrong decision will either make the student lose their interest in their designated field or perform poorly because of the institute environment. Nowadays, majority of the overseas students are applying for universities for further education in Pakistan. Unlike the students who have a Pakistani background, overseas students will have a hard time finding and then applying for the university. This platform will also assist those parents who have an uneducated background and doesn't know the technical details of finding the university which is best suited for their son/daughter.

Problem Statement

On the Internet today, an overabundance of information can be accessed, it becomes difficult for the users to process and evaluate options and make appropriate choices, especially in the case for choosing the right educational institute which is a very key factor for a student's future. The traditional method of getting recommendation is to either search global universities ranking on the internet or asking other peers for their suggestions. This is both ineffective and unreliable.

Proposed Statement

This project addresses this problem by proposing a system that provides a list of educational institutes that best suits the user preferences/needs. Different algorithms will be used in the users input such that the list is ranked according to the user's features; the university's geographical location, student's academic record/transcript, interests, budget etc. We will use different algorithms that are well-known in the Machine Learning community, which as a result can be used to produce a list of university recommendations.

Chapter 2

Literature Review

Paper Title: University Recommender System for Graduate Studies in USA By: Ramkishore Swaminathan, Joe Manley, Aditya Suresh Kumar, Swetha Krishnakumar [20]

Basic Idea

Often, the students wonder whether their profile is good enough for a certain university. This paper solves the problem by taking scrapping the required data from www.edulix.com, and a data-set containing profiles of students with admits/rejects to 45 different universities in USA was built. Based on this data set, various models were trained and a list of 10 best universities are suggested such that it maximizes the chances of a student getting an admit from that university list.

Methodologies

Data Scrapping

They Initially narrowed the list for 45 different universities. Universities with skewed data were dropped down. Then a crawler was built to get the list of students and the links to their profiles on Edulix. Once the unique set of students was identified, the data was scraped from each profile using 'Beautiful Soup'.

Data Cleansing and Transformation

About 45000 samples of raw data was obtained by as a result of scraping. Cleansing the data of had to be done, since this field was just a text box and not a selected field. A total of 1435 distinct undergraduate universities and 53 distinct majors were obtained after filtering and each of these were used as binary features.

Table 1: Statistics of the features

	Research Exp.	Industry Exp.	Intern Exp.	GRE Verbal	GRE AWE	Journal Publications	Conference Publications	CGPA
Mean	0.29	3.46	0.39	148.31	5.29	0.03	0.04	0.75
Std. Deviation	2.42	11.11	2.26	15.39	1.48	0.25	0.32	0.36
Min	0.00	0.00	0.00	0.00	1.50	0.00	0.00	0.00
Max	53.00	132.00	96.00	170.00	6.00	12.00	8.00	0.98
25%	0.00	0.00	0.00	145.00	3.00	0.00	0.00	0.70
50%	0.00	0.00	0.00	150.00	3.50	0.00	0.00	0.77
75%	0.00	0.00	0.00	154.00	4.00	0.00	0.00	0.84

Machine Learning Models Used

Three different models, ***Support Vector Machine***, ***K-Nearest Neighbors*** and ***Random Forest***, were built using a combination of all the features mentioned above, to classify a student profile to the best university that they must apply to, among the available 45 universities. Once the best university was found for the student, the 9 most similar universities in terms of the selected features was found by computing Euclidean distances to give a total of 10 universities, that the student must apply to. They split the dataset into 80:20 ratio. They used cross validation technique to avoid data overfitting as it reduces the variance by averaging over k different partitions, so the performance estimate is less sensitive to the partitioning of the data.

Result

In this work, K-Nearest Neighbor, Random Forest and Support Vector Machine were considered for recommending the 10 best universities for aspiring graduate students and their performances are summarized below:

Table 2: Accuracy of the models

Baseline	K Nearest Neighbour	Random Forest	SVM
22.2%	50.6%	50.5%	53.4%

Limitations

- Accuracy of the models predicting the data is not that great.
- This project recommends universities by taking the data of Admit/Rejects of different students from multiple universities into account only.
- Takes only 45 universities from 1047 total universities in India.

Fuzzy-Based Recommendation System for University Major Selection, By Shaima Alghamdi, Nada Alzhrani and Haneen Algethami [19]

Basic Idea

This paper focuses on the decision of choosing a university major. To make a suitable decision, a student needs an expert opinion, time, and effort. Therefore, a decision-making system should be developed in order to help prospective students to increase their educational outcome and productivity.

Methodologies

Data Collection

The first survey targeted high school students (specifically 239 prospective participants). The second survey targeted university students (specifically 392 university participants) to give their insights after spending a year in a specific major.

Overall Process

- Dataset features passed into Fuzzy Expert System (FES)
- Inference Mechanism is used
- Defuzzification is used
- Cluster Based preference is applied which puts all majors in groups, where majors in the same group are more similar than the majors in different groups

Result

Results showed that 66% were strongly pleased with the system and 54% were pleased with suggestions provided by the system.

Limitations

- Interest or preference of students is not taken into account, Dataset generated by taking the eligibility criteria for each of the major groups
- This is the first software development phase (pre-alpha version). Actual application not been implemented yet
- Focuses on university in Taif, Saudi Arabia.

Analysis and Design of Personalized Recommendation System for University Physical Education by Jun Liu, Xiaoling Wang, Xuanzheng Liu, Fan Yang

Basic Idea

This paper starts with the idea of college students' demands for P.E classes, and recommendation of an exercise is given according to the user's physicality/Fitness.

Methodologies

The system function can be divided into four parts: input module, personalized processing module, output module, and user management module, in which the personalized processing module is the core of the whole system.

Input Module

Mainly responsible for the collection and update of users' information along with the personal preference of students in sports, including the degree of endurance, preference of skill, space and number of participants etc.

Personalized processing module

This is the critical part of the system, which directly determines the performance of the system, it includes the following functional parts : confirm the identity of students in order to provide different students with different recommendations; collect students' personal preference about sports and their physical information; after the students pass the certification, the system obtains their records about personal preference and physical test; generate recommended program of physical education course according to the preference information combined with the physical conditions and the main training objectives of various courses.

Output module

This should display various types of course programs recommended to the users. The programs include recommended courses, recommended reasons, the relevant venues information, other similar courses, other sports proposals.

User management module

This is to manage the basic information for students and teachers' permissions.

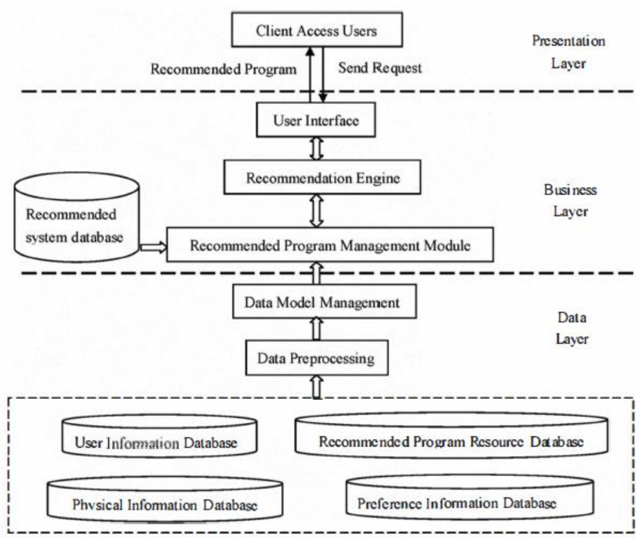


Figure 2. System application architecture diagram.

Results

The recommended process in the system designed in this paper is unidirectional. Results of the analysis will be recommended to student's; students can view the recommendation of the different results by changing the personal information input. Evaluations on the accuracy of the recommending system itself hasn't been implemented in this paper.

Limitations

- No two-way Interaction with the system through students' giving feedback and system' s modification.
- Just a concept, implementation hasn't been done yet.
- Uses Just Featured Based Collaborative Filtering (FBCF) with Pearson correlation algorithm

An Approach to a University Recommendation by Multi-criteria Collaborative Filtering and Dimensionality Reduction Techniques by: Dheeraj Bokde, Sheetal Girase, Debajyoti Mukhopadhyay.

Basic Idea

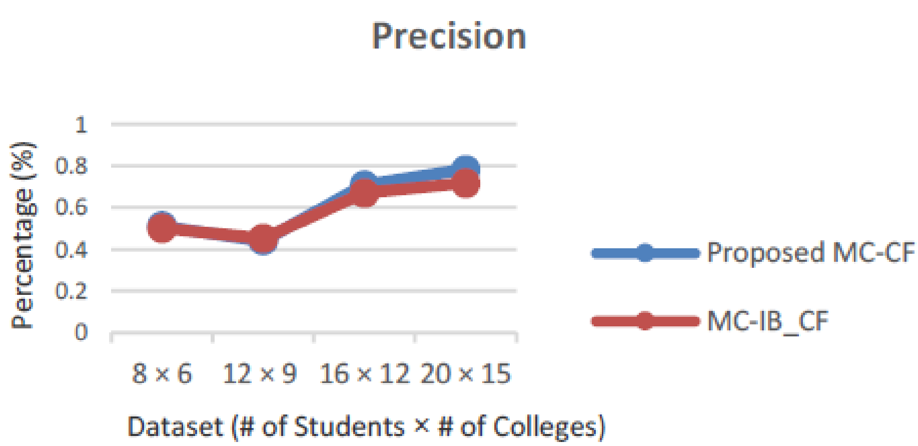
This paper proposes a University Recommendation System (URS), which provides the University or Engineering College recommendations to students, where should they apply for admission. URS makes use of Multi-Criteria Item-Based Collaborative Filtering and Dimensionality Reduction techniques to generate high quality recommendations.

Methodologies

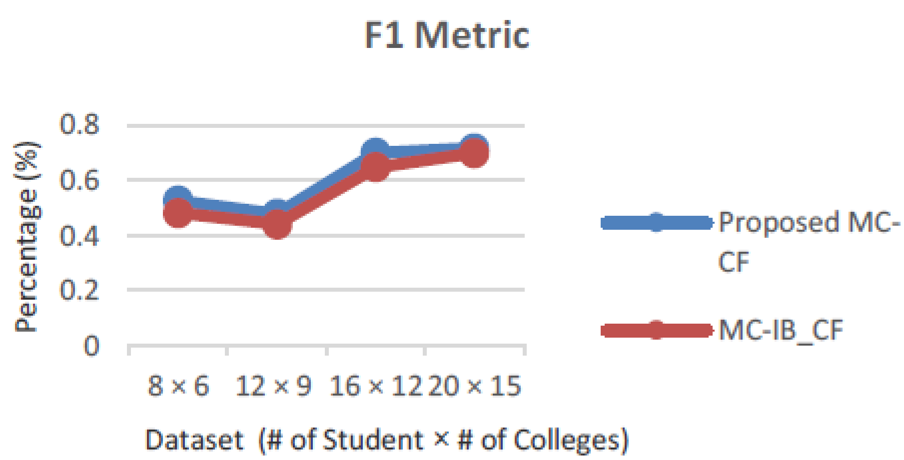
- Step 1: Generate the 3-order (student-college-criteria) tensor the from the interaction record i.e., multi-criteria ratings submitted by students for colleges.
- Step 2: HOSVD with PCA mean is applied on the 3-order tensor for dimensionality reduction to get best approximation of ratings.
- Step 3: After tensor decomposition and tensor approximation, the lower dimensional approximated data is used for similarity evaluation using cosine-similarity measure.
- Step 4: Selecting the active student and active colleges and predicting the individual criteria rating using the neighborhood formation for predicting the unknown overall rating.
- Step 5: After overall rating prediction, proposed algorithms make the predictions and list of Top-N College recommendations for the students.

Results

From the Fig. below, it can be seen that both the curve increases as the dataset size increases. However, the precision value of the proposed algorithm is larger than MC-IB_CF method. Which means more accurate results can be obtained using proposed method.



From the Fig below, it can be seen that both the curve increases as the dataset size increases. However, the F1 value of the proposed algorithm is larger than MC-IB_CF method. Which means more accurate results can be obtained using proposed method.



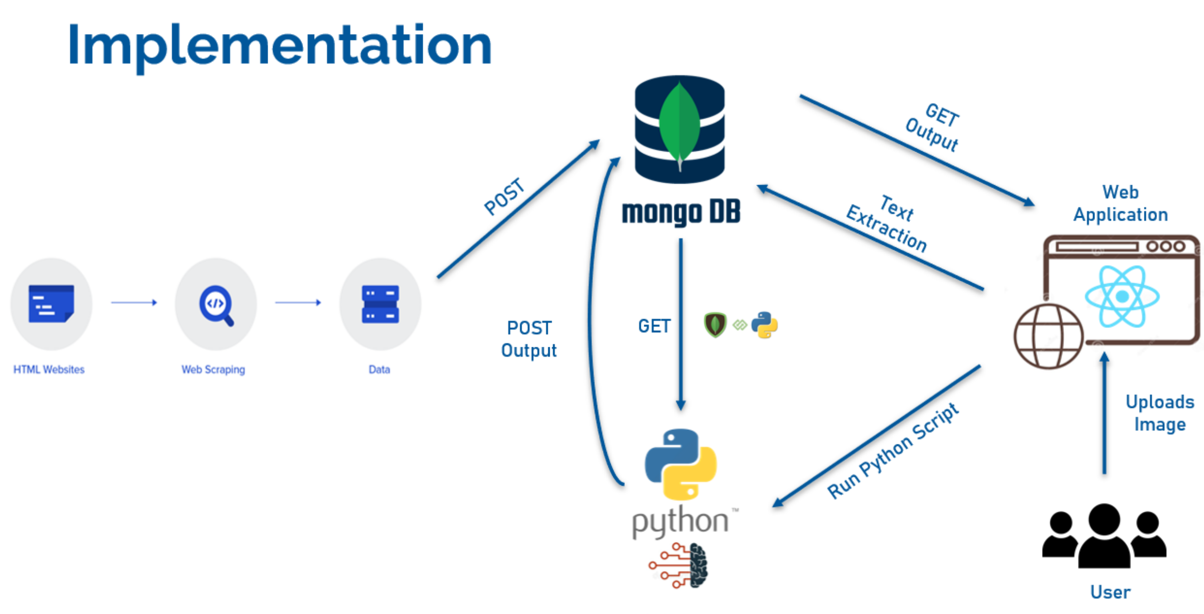
Limitations

- Affordability & the location of the institute itself isn't taken into consideration.
- Due to sparsity in rating i.e., few Students have rated the same University/College, which results slightly increase in the Execution Time.

Chapter: 3

System Analysis and Design

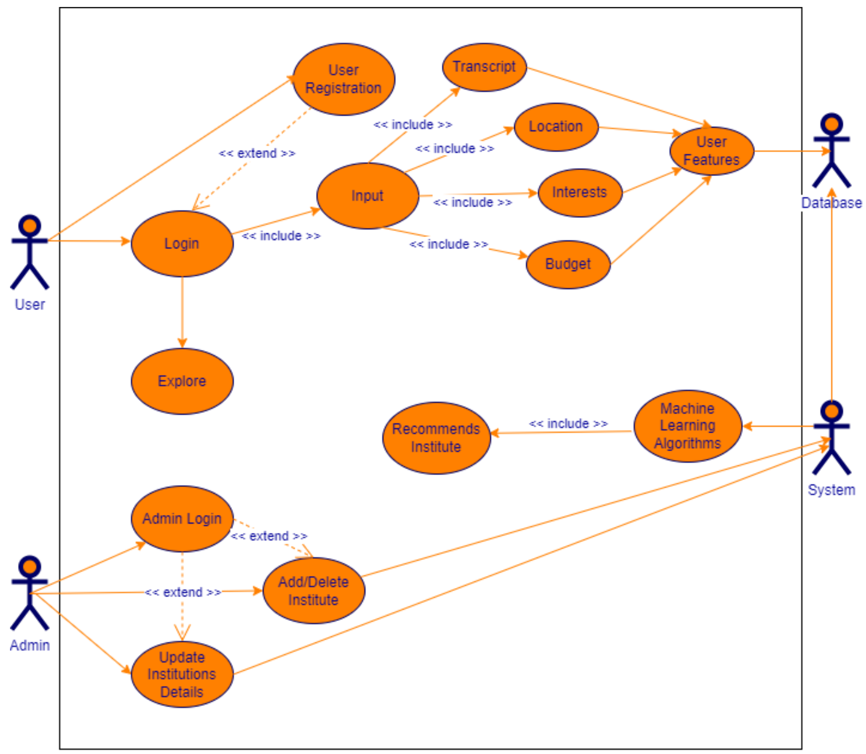
Basic Implementation



This process starts with both the user uploading his/her transcript on the REACT Application & the data collected by multiple institutes along with the dataset of different user's who are already enrolled in multiple universities is uploaded in MongoDB. Python script is run by using child_process.spawn() and the data is retrieved by using pymongo. Once the data is retrieved, different machine learning algorithm can be applied.

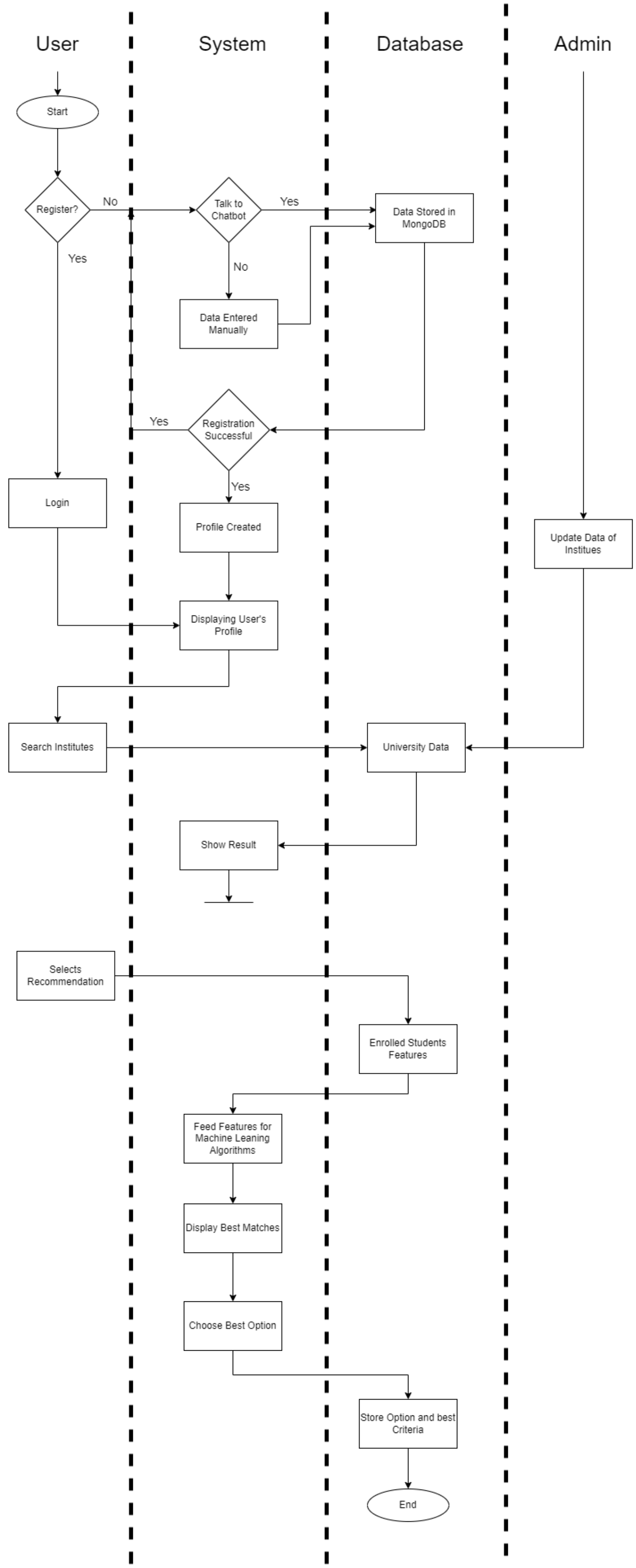
Use Case Diagram

A use case diagram represents the main use case activities and the interaction of actors and the system that is under development process. It helps to identify all the main processes of the system which are then visualized in ovals, known as a use case. A use case diagram is drawn from a scenario that explains the working of the system.

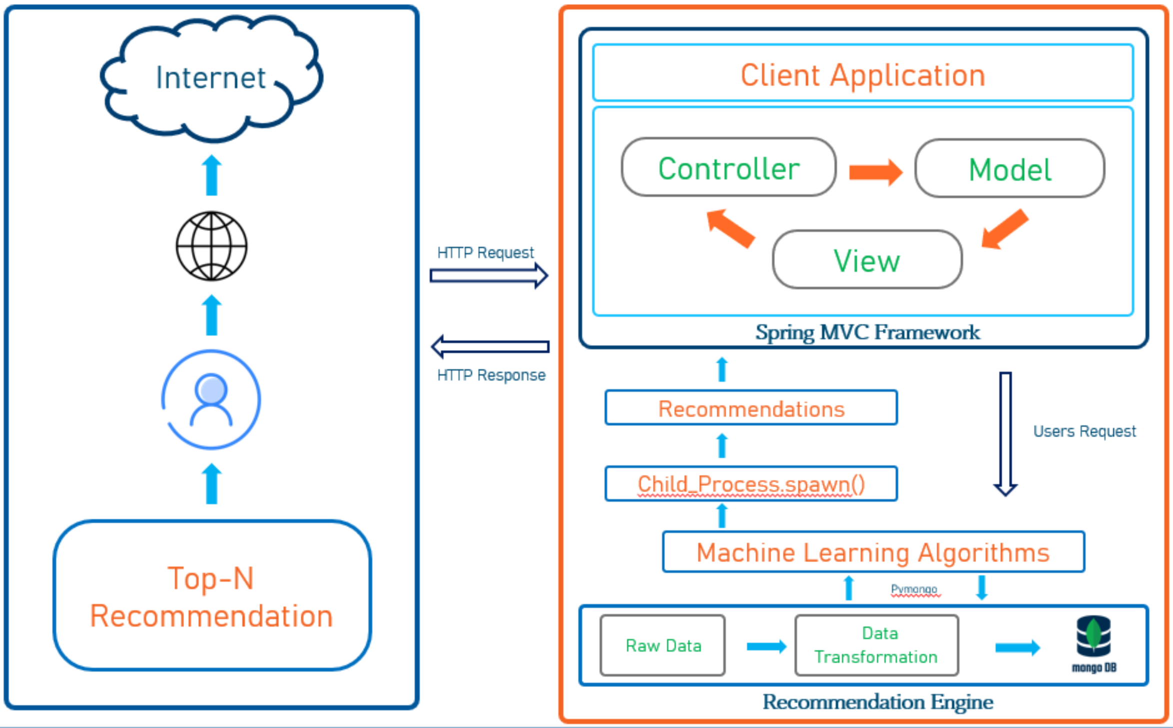


Activity Diagram

The activity diagram represents the flow of the activities in a specific order and it explains the details and conditions at every step. This UML diagram helps in understanding the flow of activities and it can help to identify those activities that can be run parallel to make an efficient system. The given diagram represents the flow of activities in Hyperdrive, it also explains which activity is initiated by which actor, since the activities are separated by swim lanes. This activity diagram is also drawn from the main scenario where the flow of activities is described.


User
System
Database
Admin
Start
Register?
No
Yes
Talk to Chatbot
Yes
No
Data Stored in MongoDB
Data Entered Manually
Registration Successful
Yes
Yes
Login
Profile Created
Update Data of Institues
Displaying User's Profile
Search Institutes
University Data
Show Result
Selects Recommendation
Enrolled Students Features
Feed Features for Machine Leaning Algorithms
Display Best Matches
Choose Best Option
Store Option and best Criteria
End

System Architecture



Bibliography

[1] J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez, Recommender systems survey, Knowledge-Based Systems, 46 (2013) 109-132.

[2] P. Resnick, H.R. Varian, Recommender systems, Communications of the ACM, 40 (1997) 56-58.

[3] Vol. 2 No. 1, April 2017, pp. 21 – 28 21 Recommendation System for Major University Determination Based on Student's Profile and Interest (e-ISSN : 2502-9401 | p-ISSN : 2503-0493) https://publikasi.dinus.ac.id/index.php/jais/article/download/1389/1057

[4] Developing and Evaluating a University Recommender System Mehdi Elahi 1*, Alain Starke1,2, Nabil El Ioini 3, Anna Alexander Lambrix 3 and Christoph Trattner https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8848746/

[5] An Approach to a University Recommendation by Multi-criteria Collaborative Filtering and Dimensionality Reduction Techniques https://www.researchgate.net/publication/301476042_An_Approach_to_a_University_Recommendation_by_Multi-criteria_Collaborative_Filtering_and_Dimensionality_Reduction_Techniques

[6] https://www.analyticsvidhya.com/blog/2021/07/recommendation-system-understanding-the-basic-concepts/

[7] https://towardsdatascience.com/recommendation-systems-explained-a42fc60591ed

[8] https://indianexpress.com/article/explained/youtube-recommendations-explained-7587523/#:~:text=YouTube's%20reccomendation%20system-,What%20decides%20the%20YouTube%20algorithm%20for%20recommendations%3F,recommend%20what%20to%20watch%20next.

[9] https://international.binus.ac.id/computer-science/2020/11/03/definition-and-history-of-recommender-systems/

[10] M. Pazzani, D. Billsus, Content-based recommendation systems, in: P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.) The Adaptive Web, Springer Berlin Heidelberg2007, pp. 325-341.

[11] M. Deshpande, G. Karypis, Item-based top-N recommendation algorithms, ACM Transactions on Information Systems (TOIS), 22 (2004) 143-177.

[12] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Item-based collaborative filtering recommendation algorithms, Proceedings of the 10th International Conference on World Wide Web, ACM, 2001, pp. 285-295.

[13] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, GroupLens: an open architecture for collaborative filtering of netnews, Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, ACM, Chapel Hill, North Carolina, USA, 1994, pp. 175-186.

[14] Q. Shambour, J. Lu, A hybrid trust-enhanced collaborative filtering recommendation approach for personalized government-to-business e-services, International Journal of Intelligent Systems, 26 (2011) 814-843.

[15] M. Nilashi, O.b. Ibrahim, N. Ithnin, Multi-criteria collaborative filtering with high accuracy using higher order singular value decomposition and Neuro-Fuzzy system, Knowledge-Based Systems, 60 (2014) 82-101.

[16]R. Burke, Hybrid web recommender systems, in: P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.) The Adaptive Web, Springer-Verlag, Berlin Heidelberg2007, pp. 377-408
[17]R. Burke, Hybrid recommender systems: survey and experiments, User Model User-Adap Inter, 12 (2002) 331-370.
[18]I. Cantador, A. Bellogín, P. Castells, A multilayer ontology-based hybrid recommendation model, AI Communications, 21 (2008) 203-210.
[19]Fuzzy-Based Recommendation System for University Major Selection, By Shaima Alghamdi, Nada Alzhrani and Haneen Algethami https://www.scitepress.org/Papers/2019/80718/80718.pdf
[20]University Recommender System for Graduate Studies in USA By: Ramkishore Swaminathan, Joe Manley, Aditya Suresh Kumar, Swetha Krishnakumar https://cseweb.ucsd.edu/classes/wi15/cse255-a/reports/fa15/026.pdf
[21] Developing and Evaluating a University Recommender System by: Mehdi Elahi 1*, Alain Starke1,2, Nabil El Ioini 3, Anna Alexander Lambrix 3 and Christoph Trattner https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8848746/
[22]Analysis and Design of Personalized Recommendation System for University Physical Education by Jun Liu, Xiaoling Wang, Xuanzheng Liu, Fan Yang