Style Guidelines for Final Year Project ReportsRecommendation System

Final Year Project Proposal

Session 2021-2022A 4th Year Student

A project submitted in partial fulfilment of the

COMSATS University Degree

Of

BSc. (Hons.)BS in Computer Science / Software Engineering (CUI)



Department of Computer Science

COMSATS University Islamabad, Lahore Campus

08 February 2021

**Project Registration**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Project ID (for office use) | | |  | | | | |
| Type (Nature of project) | | | [✓] **D**evelopment [ ] **R**esearch [ ] **R**&**D** | | | | |
| Area of specialization | | | Mobile App Development, Web App Development, Machine Learning, Natural Language Processing, Reinforcement Learning | | | | |
| **Project Group Members** | | | | | | | |
| Sr.# | Reg. # | Student Name | | CGPA | Email ID | Phone # | Signature |
| (I) | Group Leader  SP18-BCS-159 | Ibad Ahmad | | 3.06 | SP18-BCS-159@CUILAHORE.EDU.PK | 0303-2172369 |  |
| (ii) | SP18-BCS-047 | Hasseb Yaseen | | 3.27 | SP18-BCS-047@CUILAHORE.EDU.PK | 0321-8436878 |  |
| (iii) | SP18-BCS-007 | Wahaj Hafeez | | 2.7 | SP18-BCS-007@CUILAHORE.EDU.PK | 0341-2386058 |  |
| Name & Signature of Batch Advisor  (If students are eligible for FYP) | | | | Ms. Kanza Hamid (Lecturer, Computer Science) | | | |

# Plagiarism Free Certificate

This is to certify that, I am \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ S/D/o \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, group leader of FYP under registration no CIIT/\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/LHR at Computer Science Department, COMSATS Institute of Information Technology, Lahore. I declare that my FYP proposal is checked by my supervisor and the similarity index is \_\_\_\_\_\_\_\_% that is less than 20%, an acceptable limit by HEC. Report is attached herewith as Appendix A.

Date: \_\_\_\_\_\_\_\_\_\_\_\_ Name of Group Leader: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_

Name of Supervisor: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Co-Supervisor (if any):\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Designation: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Designation: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Approval of FYP Management Committee**

Committee Member 1: Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[ ] Accept [ ] \*Defer [ ] \*Reject Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\*Remarks: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Committee Member 2: Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[ ] Accept [ ] \*Defer [ ] \*Reject Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\*Remarks: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Convener: Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[ ] Accept [ ] \*Defer [ ] \*Reject Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\*Remarks: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Abstract**

Our project is a generic recommendation system which will be varying and adopting in the environment with the passage of time. In this project our aim is to provide an online recommendation system service which people can use to make their experience a lot better on contrary to their experience with some other service like Google etc. To find a correct, economical and at the same time easily accessible service is very important and needed for users but it is not an easy job today as there are whole lot of sources which are not among those which can be reliable. Today most recommendation applications are not as much intelligent and up to date as needed for users to suggest and help them in finding appropriate service that matches with the user requirement. Information mismatch have a great negative impact on such recommendation applications. To make a personalized recommendation application for providing useful and effective online services we need user reviews from online communities and up to date information from module databases.

Three models are proposed for this recommendation application. First model is the matching model which suggests the matching between user’s requirement and module databases intelligently and accurately. Second model is the quality model which is used to measure the rating index by considering the user’s opinions, ratings and reviews to help new users in finding that service. This model will be built by doing the sentiment analysis of comments about modules to generate a new rating as well as considering the rating given by previous users. Third model is the finding model which uses GPS service to find the nearest and easily accessible service. This model will be built by calculating the distance between the user location and the user destination. Finally, a mobile based and web based application is proposed to develop to demonstrate the functionality of the results of these three models.

After the static model is been trained. Our system will be able to re-train itself when the duration which we will set has been exceeded by considering the new reviews and comments or any sort or type of useful data which we can get. By following this approach our system won’t be like most of any other one that are trained ones and predicts forever but fine-tuned with the time frame and become more sophisticated and reliable

**Table of Content**

|  |  |  |
| --- | --- | --- |
| Index | Topic | Page Number |
| 1 | Introduction | 06 |
| 2 | Motivation & Scope | 07 |
| 3 | Related Work | 08 |
| 4 | System Architecture | 08 - 10 |
| 5 | Goals and Objectives | 11 |
| 6 | Individual Tasks | 11 |
| 7 | Gantt chart (Phase 1) | 12 |
| 8 | Future Work | 12 |
| 9 | Tools & Technologies | 13 |

**Introduction**

In industries like e-commerce, retail, news-group or music apps, recommendation system models are one of the most important aspects in customer retention. Presenting to the users what might interest them most is crucial. Also, identifying the most attractive content, and getting customers hooked to specific contents could result in significant revenues to the company.

Based on various data entities, including user details, interests, trending content, etc. models are built to recommend the most relevant content to customers. Companies like Spotify, Netflix, HBO use sophisticated recommendation systems for video and song recommendations. Targeted marketing is a segment of recommendation systems.

Starting with the basic idea, Recommendation system is sort of data filtering method in which basis of certain parameters you show certain data to user from database. The parameters can be rating, prices/rates, accessing comfort, nearby one etc. In simple words, Recommendation systems provide that information which directed the users to only that information which best meets their needs and preferences.

Now our goal is to create a recommendation system which will suggest its users about multiple things like doctors, restaurants, hotels etc. depending on the different parameters like ratings, remarks/feedback etc. while the parameters like affordable price, location etc. also have taken into account. On the other hand, system must be able to be adaptive with new parameter’s value which can help the system to remain up to date with the further recommendations.

Now if we talk about the basic parameters, these are as following

* Rating
* Reviews
* Nearby
* Price

**Motivation and Scope**

Since we are living in the 3rd word country and are among a developing ones in Asia we don’t have recourses that can cover up the mass requirements. There are hundreds of thousands of cases weekly that are unable to get even first aid timely especially those who are lying in or near to rural areas. Pakistan’s population which belongs to this section is reported to be 63.09% in 2019 which is a shocking stat itself.

One of the many examples we can site here is of **Muniba Mazari** also known as the Iron Lady of Pakistan. She is a Pakistani activist, anchor artist, model, singer and motivational speaker. On 27 February 2008, Muniba and her husband were travelling from [Quetta](https://en.wikipedia.org/wiki/Quetta) to [Rahim Yar Khan](https://en.wikipedia.org/wiki/Rahim_Yar_Khan_District). Their car met with an accident, in which she sustained several major injuries, including broken bones in her arm (both [radius](https://en.wikipedia.org/wiki/Radius_(bone)) and [ulna](https://en.wikipedia.org/wiki/Ulna)), rib-cage, shoulder blade, collarbone and spine. Her [lungs](https://en.wikipedia.org/wiki/Lungs) and [liver](https://en.wikipedia.org/wiki/Liver) were also deeply cut. Moreover, her entire lower body was left paralyzed. She was taken to a nearby hospital, which was ill-equipped to deal with such a severe case. She was then taken to her native hospital but result didn’t change as there were no doctors or equipment’s to handle the patient either. She was then moved to a hospital in [Rahim Yar Khan](https://en.wikipedia.org/wiki/Rahim_Yar_Khan_District), and eventually, she was admitted to the [Agha Khan Hospital](https://en.wikipedia.org/wiki/Aga_Khan_University_Hospital,_Karachi), [Karachi](https://en.wikipedia.org/wiki/Karachi). Post-surgery, she was left bed-ridden for two years.

The idea to cite the above example is there are several cases of these context where the population is unable to find the emergency service just because of the unawareness and no helping hands that can lead them to desired output. Our aim is same, to provide that sort of application which will be able to guide with best possible source of help that user can imagine or desire at runtime.

This application will cover most of the modules that we counter in normal life ranging from Hotels, Libraries, Malls, doctors etc. Furthermore, user can apply filters as per their use. In future, more modules can be added to this system as user requirements increase and so do filters.

**Related Work**

* A few years ago, Netflix organized a challenge “Netflix Prize”, where they invited people to build a better recommender system than what they had in return for prize money.
* There is an article on "Analytics Steps" named "What Are Recommendation Systems in Machine Learning?” This article helps to understand the types of Recommendation systems in the market. On the other hand, according to some articles like 1 on "Medium.com" named "Introduction to recommender systems," there are 2 major approaches i.e. Collaborative & Content base, and explained in detail in that article. Collaborative is the one under which our project lies. If we talk about the related work which falls under the category of the Collaborative Filtering type, there are many like Food Panda to recommend good food nearby, similarly booking.com recommend best hotels to its user according to budget, ratings, etc. Each of that platform is single task specific recommender but our goal is to develop single platform for all such type of recommendations.
* Our Supervisor Ms. Kanza Hamid has already supervised a similar type of project named ‘Doctor in Hand’ which was a static application that recommends best doctors in town along the appointment booking section for the user.

**System Architecture**

**Proposed System**

We have divided the proposed system into 3 phases as follows:

1. **Data Collection**

This phase consists of collecting information about different modules like doctors, restaurant, hotel etc. by taking user’s comments and ratings about modules. For completing this part of the system, we have scraped the data from web by doing programming in Python. After collecting the data, data cleaning operations are performed to remove the unwanted or corrupted data.

1. **Recommendation System**

The data collected in phase one is passed to this phase. This phase of system is further divided into three models that are explained as:

**2.2.1 Relevance Model**

The relevance model finds the matching of user requirement with the different models and then compare the requirement with the dataset of that model. It will consist of search field to enter text as well as the option to choose filters. When the user search for anything, first matching user requirement by model and then matching between user requirement and the database will be performing to get relevant profiles so in the next model accurate recommendation will generate.

**2.2.2 Quality Model**

This is the core model for generating recommendations. Natural language processing techniques such as tokenization, stemming, removing stop words and bag of words will be performing on raw text comments to convert them in a meaningful form which can be understandable by machine to learn. Machine learning technique such as sentiment analysis will be carried out after passing the comments given by users about modules from natural language processing pipeline. Sentiment analysis will perform to know the polarity of these comments to check whether it is positive, negative or neutral to use this sentiment score for further processing.

**2.2.3 Find nearby Model**

In this model, distance between the user location and the object location will be calculated to find the nearby object. User location will get by enabling his/her mobile GPS whereas the object location will be get from Google Map. Nearby doctors, restaurant, hotels etc. by showing the distance in km or meters will be display to user.

After the overall analysis of module’s recommendation priority using these three models, a system rating will be generated that will be used for recommending the best object from a list of objects.

1. **Reinforcement Model**

As we discuss that system must be adaptive so here comes reinforcement model in action. After a certain period of time our model will retrained on the data collected from new remarks and generate new ratings which will help our system to be more reliable and consistent.

**Proposed Algorithm**

1. Pass the comments to Natural Language Processing Pipeline which consists of

i. Text Cleaning

ii. Tokenization

iii. Stemming

iv. Removing Stop words

2. Feature Extraction using Count Vectorizer.

3. Sentiment Analysis using Machine Learning Algorithm Deep Neural Network to classify the comment into positive, negative or neutral.

4. Generating System rating using the formula defined above.

5. Displayed the Best recommendation by sorting the rating in decreasing order.

6. Apply reinforcement model on new feedback to generate new rating and update the previous ones

**Goals and Objectives**

* To make it easy and effective for users to get best recommendations nearest to their location and easily accessible.
* To develop such application that will automate the function of finding the appropriate recommendation according to user requirement, thus bring easiness, save time and efforts needed to find a best suitable suggestion.
* To overcome the drawbacks of traditional approaches for finding a thing by developing such recommendation system that will uses natural language processing and machine earning techniques to produce a recommendation list.
* To develop such intelligent recommendation system that performs accurately and efficiently and help users to know about the authentic, best, affordable, nearest doctor, restaurants, hotels etc. whom which they can trust or prefer.
* To create a model, that will vary time to time in order to acquire excellence in further recommendations.

**Individual Tasks**

Data collection is done by all members which includes finding suitable sites for data scraping and creating a scraper. After that data refinement will be done by 1 member while the other will work on Database ERD. Moving towards the next phase, each of the members will work on the proposed recommendation model which includes the NLP task each, will contribute their efforts to achieve the best models possible. In the development phase of app and web, each member will work to develop the platform in which he is best like one on native the other on react, and the one on the backend, etc. The testing phase will be done by each group member in which each member will try to test the developed product of others. Last but not the least, each member will contribute to report writing according to the availability of members.

**Gantt chart**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Task Description** | **Months** | | | | | | | | | | | | | | | |
| **FEB** | | **MAR** | | | | **APR** | | | | **MAY** | | | | **JUN** | |
| **Information Gathering** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Data Scrapping** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Data analysis and refinement** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Database Development** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Apply NLP, ML Techniques** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Report Writing** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Future Work**

To make this application more useful and attractive for the people to use it and get benefit from it, following are some points that are in our mind to include in this application in future:

* Integrate more modules over the time according to their need.
* Add the functionality of online booking to save the time and the effort required to get appointment or booking by visiting.
* Make blog within application where ranked publish their experience and reviews to aware the people about the latest insights

**Tools and Technologies**

* **Python**

It will be used for training, training phase of the model, and for data scrapping using its libraries like Tensorflow, Keras, Pandas, Numpy, beautifulSoup, Sciekit-Learn etc.

* **Google Colab or Kaggle Kernels**

It provides a GPU support for models training and testing freely. Will be a massive asset since the second and third model require the Deep Learning which takes hours and days depends upon the problem statement on CPU’S. So the GPU support is mandatory

* **JavaScript**

JavaScript, often abbreviated as JS, is a programming language that conforms to the ECMAScript specification. JavaScript is high-level, often just-in-time compiled, and multi-paradigm. It has curly-bracket syntax, dynamic typing, prototype-based object-orientation, and first-class functions. For web and native app development, we will use JavaScript libraries like node js, express js, react native and the list goes on.

Appendix A

*Include here the 1st page of Turn Tin Report*

Every supervisor has his/her own Turn tin account. If not then supervisors are requested to get the account from Library as soon as possible.