Recommendation System For Tourist Destinations Using NLP

Project Team

Zohaib-Khan p19-0095
Ahmed-Ali p19-0042

Session 2019-2023

Supervised by

Ms Salma-Kalsoom

Co-Supervised by

Mr Zeeshan Khan

Department of Computer Science

National University of Computer and Emerging Sciences
Peshawar, Pakistan

June, 2023

Student's Declaration

We declare that this project titled "*Recommendation System For Tourist Destinations Using NLP*", submitted as requirement for the award of degree of Bachelors in Computer Science, does not contain any material previously submitted for a degree in any university; and that to the best of our knowledge, it does not contain any materials previously published or written by another person except where due reference is made in the text.

We understand that the management of Department of Computer Science, National University of Computer and Emerging Sciences, has a zero tolerance policy towards plagiarism. Therefore, We, as authors of the above-mentioned thesis, solemnly declare that no portion of our thesis has been plagiarized and any material used in the thesis from other sources is properly referenced.

We further understand that if we are found guilty of any form of plagiarism in the thesis work even after graduation, the University reserves the right to revoke our BS degree.

Zohaib-Khan Signature: _____

Ahmed-Ali Signature: _____

_____

Verified by Plagiarism Cell Officer

Dated:

Certificate of Approval

The Department of Computer Science, National University of Computer and Emerging Sciences, accepts this thesis titled *Recommendation System For Tourist Destinations Using NLP*, submitted by Zohaib-Khan (p19-0095), and Ahmed-Ali (p19-0042), in its current form, and it is satisfying the dissertation requirements for the award of Bachelors Degree in Computer Science.

Supervisor

Ms Salma-Kalsoom Signature: _____

Co-Supervisor

Mr Zeeshan Khan Signature: _____

_____

FYP Coordinator
National University of Computer and Emerging Sciences, Peshawar

_____

Dr. Hafeez ur Rehman

HoD of Department of Computer Science
National University of Computer and Emerging Sciences

Acknowledgements

Zohaib-Khan

Ahmed-Ali

Abstract

Having knowledge of tourist places for tourists is a difficult task in unfamiliar countries. Many tourist recommender system only consider broad POI categories and do not align well with users' preferences and other locational constraints.Mostly tourist does not give feedback on the places they are visiting which makes it more difficult to recommend different good places to visit.We don't have sufficient ratings of the places given by the tourists.They are not actually intelligent systems only they are giving results from the databases based on different categories.Our tourist places recommendation algorithm is modelled in natural language processing, coupled with an TF-IDF algorithm, Semantic Analysis, Sentimental Analysis, syntactic Analysis.

Contents

List of Figures

List of Tables

Chapter 1

Preliminaries and Introduction

During the past few years, with the explosion of information being serve as online, it is difficult for users finding relevent information according to their interest.Their is a great impact on economy by having tourism opportunities.In many countries different types of scams happening with the tourists so by having an application which recommend places according to their interest, there is a lesser chance of having scams to tourists.By having tourism it shows a softer image of your country.It helps to grow your economy and helps in reducing unemployment. The primary goals of recommendation systems are to encourage demand as well as to actively attract users.Since one of the main aspects of recommendation systems is pro-activity, they will suggest information to the user that may be of interest without the user having to look for it explicitly and in accordance with his or her predictions.These recommendations are based on user's past activity, attributes, contextual information and in similar users based information. [**?**]

1.1 What is a Recommendation System?

A recommender system is a system which predicts ratings a user might give to a specific item. These predictions will then be ranked and returned back to the user. They're used by various large name companies like Google, Instagram, 2 Spotify, Amazon, Reddit, Netflix etc. Most Recommendation systems can be classified in three different groups:

Content-based, Collaborative Filtering and Hybrids.

1.1.1 Collaborative Filtering Systems

A collaborative filtering recommender system displays recommendation based on preferences of similar users.The results are based on feedback from users who are similar to target user instead of on the target users on fast preferences.The accuracy of collaborative filtering method depends on the number of items which can be associated with certain users.

1.1.2 Content-based Systems

Content-based filtering uses item-features to recommend other items similar to what the users likes, based on their previous actions explicit feedback.The model should recommend items relevent to this users to do so you must first pick a similarity matrix. Then you must setup a system to score each candidate item according to this similarity matrix.

1.1.3 Hybrid recommendation

Hybrid recommendation systems are essentially combinations of the previous two approaches. For Example, they can use collaborative filtering and content-based information together to suggest a broader range of items accurately. The complexity of these hybrid systems tends to make them more expensive for the content producer to implement.Hybrid recommendation systems are essentially combinations of the previous two approaches. For Example, they can use collaborative filtering and content-based information together to suggest a broader range of items accurately. The complexity of these hybrid systems tends to make them more expensive for the content producer to implement.

1.2 Recommendation system history

Recommender Systems (RS) 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. 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 Taperstry, the first commercial RS. Another RS implementation for helping people find their preferred articles was launched in the early 1990s by GroupLens. Further development of RS in the late 1990s was the implementation of Amazon Collaborative Filtering, one of the most widely known RS technologies. Since this era, RS based on Collaborative Filtering has become very popular and has been implemented by many e-commerce and online systems. Many toolboxes for RS have also been developed. The success story of Amazon also gave rise to the development of many RS algorithms known as hybrid approaches, which combine multiple approaches. Following the successful era at the end of 1990s, the industry offered generous funding to implement RS research. The most popular competition in RS was held by Netflix, a provider of internet streaming media. They launched the Netflix prize 1 in 2006 and give 1 million US Dollars to the winner of the combinations who provided the best RS movie recommend. They announces the winning team in 2009. In 2010, YouTube also implemented an RS on its website

1.3 Motivation

There are many tourists that are not aware of tourist spots and just put their money on risk in visiting those places so we consider it's a major problem and comes with intelligent RS using NLP.This will make ease for tourists in searching of places in Pakistan.

1.4 Problem statement

There is a major lack of recommendation system in today's tourism applications.In most of the applications they are generally designed for not recommendation but just giving a relevent results from the database. In Pakistan there is no smartness in tourism applications.

Chapter 2

Review of Literature

2.1 Overview

Our primary goal is to provide user a good recommendations according to their interests, a user will provide us the context , and our work is to extract the meaningful keywords,from that particular context.for example ,if a user enter text like mountains with crystal clear blue water , so our work starts from collecting the keywords(meaningful) such as from this , we get keywords like crystal-clear ,water , mountains .

2.1.1 Main Features

Our project has three main features **application, extraction of keywords , matching of keywords from labeled images**

2.2 Recommendation filtering techniques

Content-based technique is a domain-dependent algorithm and it emphasizes more on the analysis of the attributes of items in order to generate predictions. CBF uses different types of models to find similarity between documents in order to generate meaningful recommendations. It could use Vector Space Model such as Term Frequency Inverse

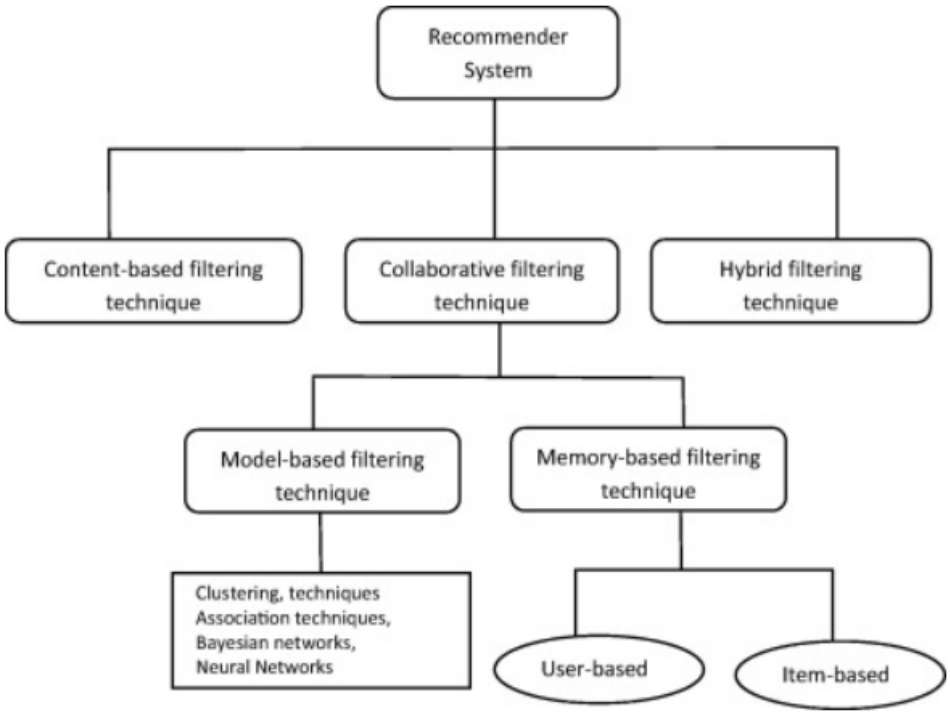

Figure 2.1: Recommendation filtering techniques

Document Frequency (TF/IDF)

2.2.1 Content-based filtering

Content-based technique is a domain-dependent algorithm and it emphasizes more on the analysis of the attributes of items in order to generate predictions. Content-based filtering technique does not need the profile of other users since they do not influence recommendation. The major disadvantage of this technique is the need to have an in-depth knowledge and description of the features of the items in the profile.

2.2.2 Collaborative filtering

Collaborative filtering technique works by building a database (user-item matrix) of preferences for items by users. It then matches users with relevant interest and preferences by calculating similarities between their profiles to make recommendations Recommenda-

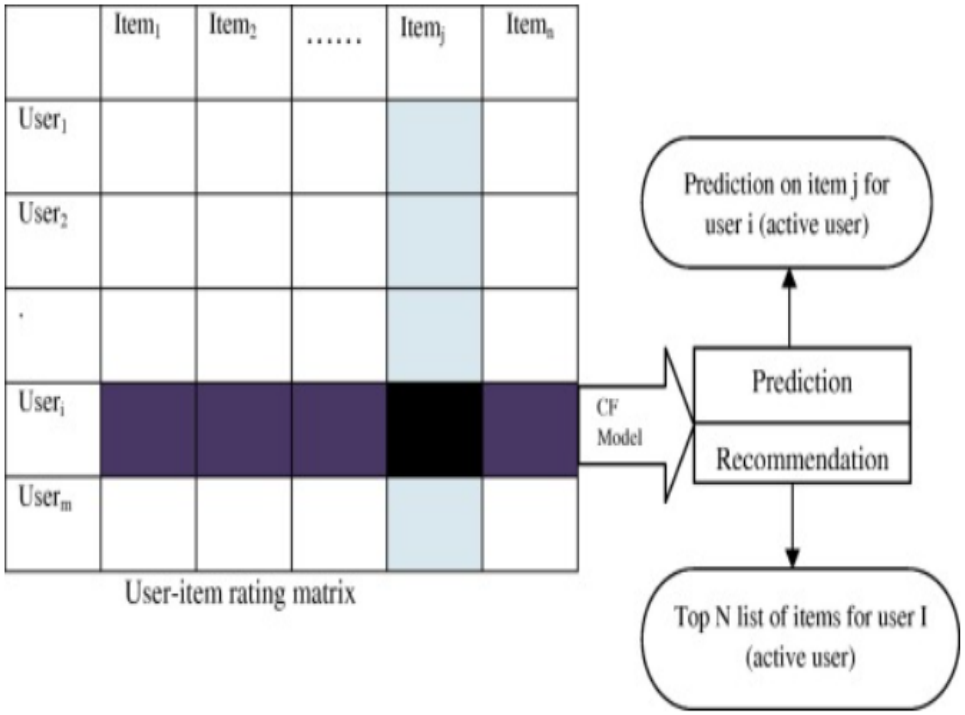

Figure 2.2: Collaborative filtering

tions that are produced by CF can be of either prediction or recommendation. Problems in Collaborative filtering are **Cold-start problem, Data sparsity problem(lack of enough information) , Scalability**

2.2.3 Methods to solve Scalability Problem

Methods used for solving scalability problem and speeding up recommendation generation are based on Dimensionality reduction techniques, such as Singular Value Decomposition (SVD) method, which has the ability to produce reliable and efficient recommendations.

2.2.4 Hybrid filtering

The idea behind hybrid techniques is that a combination of algorithms will provide more accurate and effective recommendations than a single algorithm as the disadvantages of

one algorithm can be overcome by another algorithm. The combination of approaches can be done in any of the following ways: separate implementation of algorithms and combining the result, utilizing some content-based filtering in collaborative approach, utilizing some collaborative filtering in content-based approach, creating a unified recommendation system that brings together both approaches.

Chapter 3

Review of Literature

In this chapter, we will analyze our project in terms of Project manage- ment, System analysis, and design. Project management includes different stages such as planning, analysis, design, deployment, and maintenance. In system analysis, we have made different use cases for our project so that we can design our project based on these use cases. A flowchart and activity diagram are given so that a proper workflow should be followed to assemble this project. We will also discuss our Proposed Methodology in this section as well

3.1 Use Case Diagram

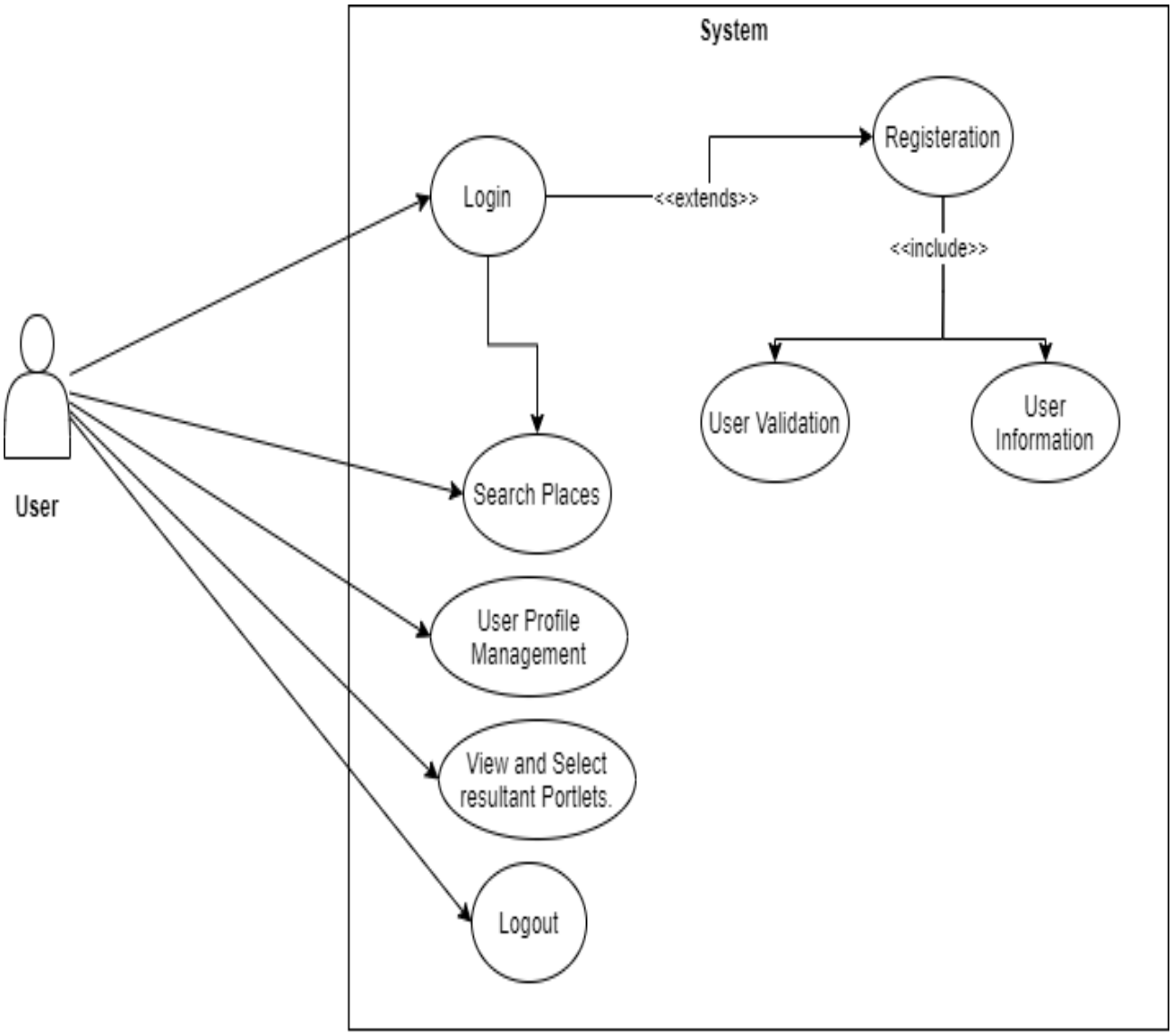

Figure 3.1: Use Case Diagram

3.2 Activity Diagram

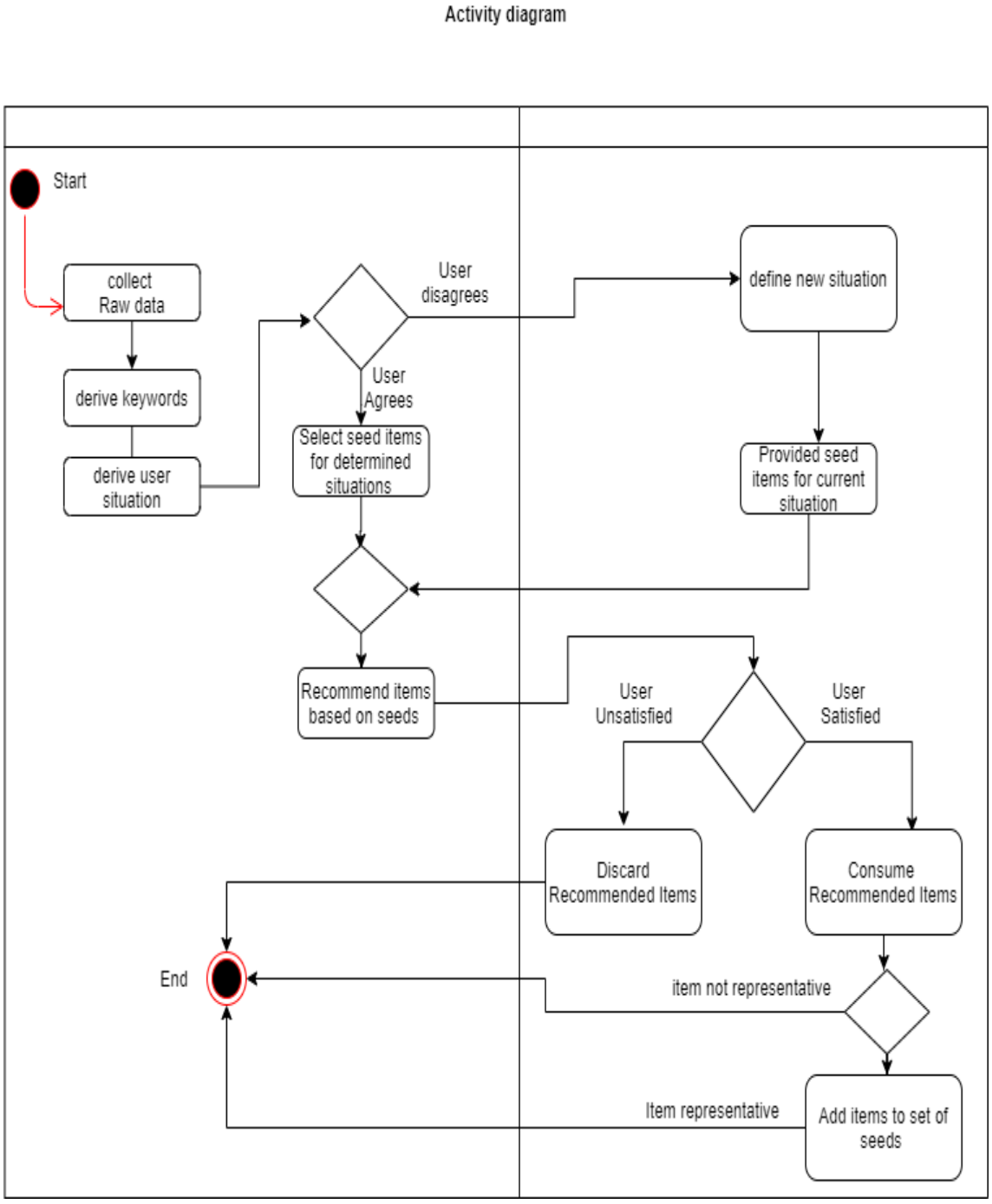

Figure 3.2: Activity Diagram

3.3 Sequence Diagram

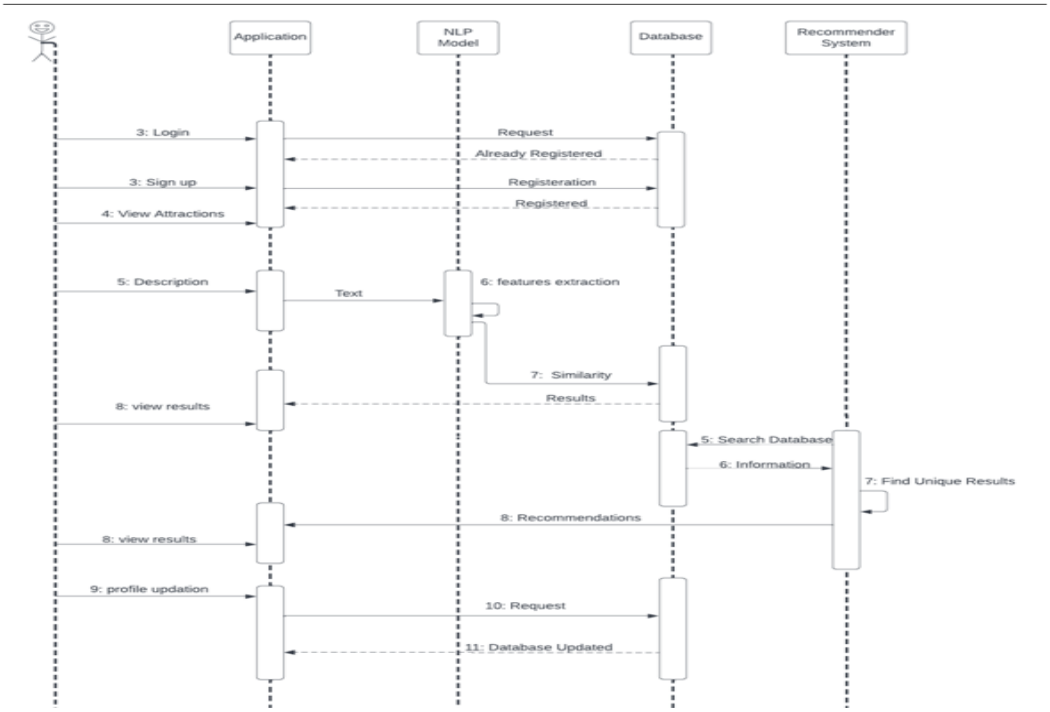

Figure 3.3: Sequence Diagram

Chapter 4

Proposed Methodology

4.1 Introduction

In this chapter, we will discuss the detailed methodology that will be carried out in our project.

. Labeling of local dataset(Images of different places KP only)

. data Acquisition

. Pre Processsing (NLTK Library (StemmingwordNet lemmitizer) ,Word Tokenization, StopWords(use to exclude those words which are irrelevant))

. Feature Extraction(Tf-Idf),cosine similarities

. spacy library(more accurate than NLTK library)

. word2vec (genism package)

. BERT

. Torch

. sklearn

. numpy

.

Chapter 5

Result Analysis

5.1 Outcomes

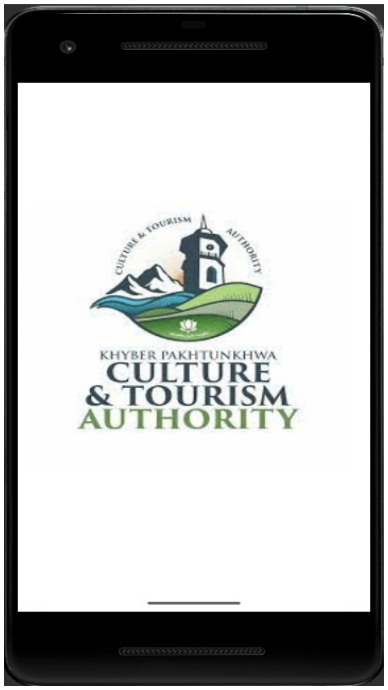

Figure 5.1: shows the pose detection and estimation.

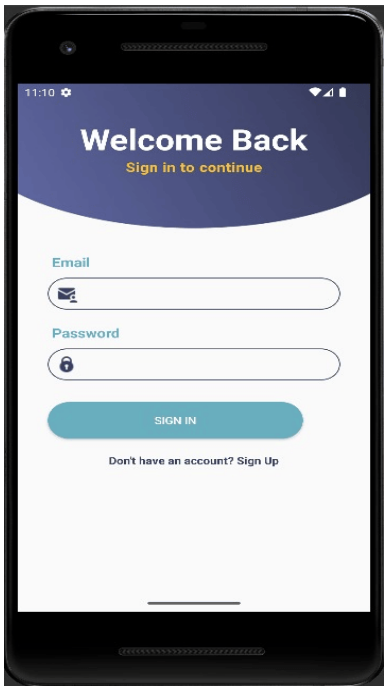

Figure 5.2: shows the pose detection and estimation.

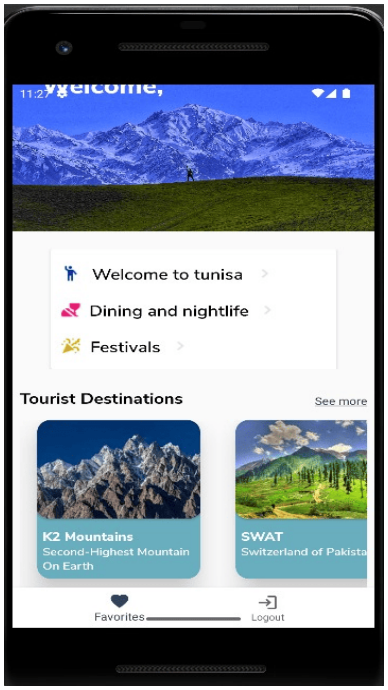

Figure 5.3: shows the pose detection and estimation.

	ID.1	City	sentence
2804	2805	Vienna	This is the perfect activity for a hot summer day! Combining an easy hike with a bath in the crystal clear mountain river!\n\nThe trail runs basically along the river in a valley, offering fantastic views to the surrounding mountain scenery. It is for sure one of the most magnificent gorges of the Eastern Alps. The water is crystal clear and we can hear its sound all the time, which has a very relaxing effect on body and soul.\n\nThe path is a mountain trail and narrow, rocky and rooted most...
470	471	Athens	(Discounted price for limited time)\nDiscover breathtaking views in this specially designed hike!\n\nVouliagmeni combines turquoise beaches with majestic mountain views. Semi dry plants, rare birds & mammals even a thermal lake can be found here. Among them lie a network of very old footpaths used by the first shepherds to move their herds. \n\n>We will start our trek just outside the lake of Vouliagmeni & we will gradually hike on a nearby hill\n>The whole trek offers great 360o panoramic ...
2168	2169	Prague	We will go to one of the most unique and magical natural places of the country - Czech Switzerland. On our trip we will see amazing Elbe sandstones, which are a magnificent corner of preserved nature, ride a gondola along a picturesque river with steep rocky shores, and then go up to the ancient Falcon Nest hotel, built high in the mountains.

Figure 5.4: shows the pose detection and estimation.

```
[10] query_sentence
0s

'a mountains with a beautiful lakes'
```

Figure 5.5: shows the pose detection and estimation.

Chapter 6

Problem Discussion

6.1 Drawbacks

In most of the recommendation system , they give the results from databases , they don't use any machine learning model for for smart recommendation. Cold start problem.

6.2 Proposed Solution

We are using different models of machine learning , the models with higher accuracy , we combine their results and give most accurate results according to the choice of the user. we are eliminating the cold start problem , by using a hybrid approach.

Chapter 7

Conclusions and Future work

7.1 Conclusions

we process the given sentences , find keywords ,match them with the label of images , give recommendations accordingly.

7.2 Future work

classify the data-set according to our need, find models which gives highest accuracy , then use that model in our recommender system , introduces new functionalities like giving weather forcast etc. Built an application .