East West University

Capstone 400A Project Report

Submitted To:
Md. Mostofa Kamal Rasel
Associate Professor,
Department of Computer Science and Engineering
East West University, Dhaka

Submitted By:

Name	ID Number
Farhana Jaman	2020-1-60-006
Noshin Tasnim Khan	2020-1-60-221
Fima Faria Lisa	2020-1-60-026
Ayesha Chowdhury	2020-1-60-003

Department of Computer Science and Engineering
East West University, Dhaka

"Efficient Tourism Route Planning in Bangladesh: A Time-Conscious Tourist Places Recommendation System"

1. Introduction:

Tourism is a significant industry worldwide, with travellers often seeking personalised experiences according to their preferences. For planning purposes, they rely on guidebooks, word of mouth recommendations or online search engines, which fail in delivering proper suggestions according to their preferences and are very time consuming. Sometimes tourists may consult travel agencies to plan their trip. Tourists have to plan their trip according to agencies' available packages. Each tourist's interest and need are not considered when a trip is planned by travel agencies. People in agencies convince tourists for their packages even if tourists are not interested. Generalised packages are offered by travel agencies and do not give freedom to tourists to choose their hotels, restaurants according to their choice and need[1]. Recognizing these challenges, the Recommendation system is a good solution to meet the needs of travellers according to their preferences. However, existing tourist recommendation systems often fail to recognize the duration of travel time of travellers. To overcome this challenge, this project will create a tourist place recommendation and route planning system that takes into account the time constraints of travellers. Here, travellers can specify their available time frame, such as between 8 am to 12 pm and the system will generate the most interesting places to visit for travellers. It also shows the travel route among the present and all the tourist places that can be visited within the time frame. Furthermore, we will use a real dataset of Bangladesh. In this dataset, every interesting place in every upazila (sub-district) of Bangladesh has been stored. We created this dataset using google place api and using the website of bangladesh where we found the names of all upazila of bangladesh.We have already written a paper about this dataset.

2. Literature review:

Prof. Sharikant et al. [1] proposed a personalised recommendation system which recommends places based on user interest,current location and range of distance. It uses content based filters to find similar interest among user profiles. Also, calculate euclidean distance and use KNN algorithm to recommend the places. Here, the apriori algorithm is used for classification. This system also filters out visited places, providing information like reviews and seasonal recommendations.

Mishal et al. [2] introduced a personalised and budget friendly tour spot recommendation system based on content-based filtering, which helps to enhance the tourist experience by recommending the best picnic spots according to all user's preferences, including budgets, interests, destination country, type of places, safety, transportation and climate. This system compares the features of different spots and takes into account the user's previous history to recommend similar tour spots by using content based filtering methods and predict the user's preferences accurately and suggest the best picnic spots.

In the paper [3] "Measuring Tourist Experience in Semarang City through an Advanced Recommendation System," which is written by authors Rudi Sutomo and Daffa Kaisha Pratama from Universitas Multimedia Nusantara present an innovative approach to enhance tourism experiences in the City of Semarang , Indonesia. They develop an advanced recommendation system which integrates content-based and collaborative filtering techniques to offer personalised tourist suggestions. The paper also demonstrates the system's efficacy in simplifying travel decisions and uncovering Semarang's hidden treasures. Overall, it fills a research gap and also contributes to the field of tourism recommendation systems.

In The paper[4] "Tourist Place Recommendation System Using Machine Learning" by Chinmay Natu, Indrajeet Patil, Yogesh Ghaste, Prathamesh Navangul, and Prof. Vijay D Chougule discovers how machine learning can enhance tourism recommendation system. It contributes to a broader field of research aiming to improve tourist experiences by leveraging user-generated content, personalised recommendations, and interactive interfaces. And also highlight the importance of collaborative, location-based, and image-based recommendation systems in the tourism system. Generally, the paper focused on advancing tourism recommendation technologies.

Title	Activity	Algorithm	Limitations

Traveler's Recommendation System Using Data Mining Techniques.	Recommend considering user's interest, current location and range of distance.Filter out visited places.	Content Based Filtering, KNN algorithm and Apriori algorithm	Used Euclidean distance to calculate distance so exact distance cannot be found. Does Not consider travellers travel time, budget, climate, safety etc.
Tour Spot Recommendation System via Content-Based Filtering	Enhance the tourist experience by recommending the best picnic spots according to all user's preferences, including budgets, interests, destination country, type of places, safety, transportation and climate.	Content Based Filtering	Development of ontology can improve accuracy. Does not consider travellers travel time.
Measuring Tourist Experience in Semarang City through an Advanced Recommendation System	Enhance tourism experiences in the City of Semarang , Indonesia and demonstrates the system's efficacy in simplifying travel decisions and uncovering Semarang's hidden treasures	content-based and collaborative filtering	dataset which is obtained from Kaggle, which may not fully represent all aspects of tourism in Semarang City and also The content-based and collaborative filtering models which is used in the system may oversimplify tourist preferences and behaviours, leading to less accurate recommendations.
Tourist Place Recommendation System Using Machine Learning	contributes to a broader field of research aiming to improve tourist experiences by leveraging user-generated content,	collaborative, location-based, and image recognition algorithms	Limited data availability may affect recommendation accuracy and also Recommendations may be biassed based on the dataset used

	personalised recommendations, and interactive interfaces		

3. Motivation

The motivation of this project is to overcome the limitations of the traditional recommendation system for travels. These systems often overlook the time constraints faced by the travellers. Travellers often have specific time slots available to explore places due to flight schedules or other reasons. They may have time only in the morning or afternoon to travel. For example, they can have time between 8 am to 12 pm to visit places. So, to overcome this problem, we get motivated to develop a recommendation system that considers time constraints of travellers. Our system will recommend the best places to visit by considering the traverler's available time frame considering the current location. Also, It will show the travel route among the present location and all the tourist places that can be visited within the time frame. This will help travellers make the most of their limited time having a fulfilling travel experience.

Furthermore, there is a lack of a recommendation system on the dataset of Bangladesh, which includes information about tourist attractions in every upazila (sub-district) of Bangladesh. Also, a dataset of tourist places in Bangladesh is not available. So, we have collected data using google place api and created a dataset where tourist attractions of every upazila in bangladesh is stored. So this dataset will be used for the recommendation system so the system could provide accurate and relevant recommendations considering the time constraint and tourists can explore our beautiful Bangladesh to make the most of their limited time.

4. Problem Statement

The problem is, in the dataset, there is no information about time duration to travel between tourist attractions and places. So how can I recommend tourist places to travellers within a time frame?

Solution: Firstly, we can use google map api to get to know about the duration of time to travel between one place to another. But to use google map api we have to give them credit card information and possibly they will charge money after some limited uses. Another solution is, in the

dataset, there is information about latitude and longitude of the tourist places. So, we can possibly calculate the distance among tourist places and distance from the present location. After calculating the distance, we can measure time that will take to travel.

5. Objective

i) Recommend tourist places and plan routes considering the time frame where starting time, departure time will be given to do recommendations.

ii) Developing a tourist recommendation system where tourist places in every upazila of Bangladesh can be found.

6. Proposed Work

Our objective is to develop a recommendation system that recommends tourist places to which a traveller can visit within a specified time constraint, such as from 10 am to 4 pm, in a particular Upazila, utilising latitude and longitude coordinates. Additionally, we will calculate the distance from the latitude and longitude coordinates.Manhattan Distance helps recommend tourist spots by measuring the time between the traveller's location and potential destinations within an Upazilla.After calculating this time, the system will prioritise nearby attractive places which help travellers to make the most of their time constraints and explore efficiently within the specified timeframe.

6.1. Datasets:

Dataset containing details about famous places in each Upazila, including attributes such as ID, search_key, category, place-name, title, address, latitude, and longitude, rating, rating counts.

6.2. Design/Framework:

- Python programming language for implementation.
- Google Maps API for geolocation services.

A proposed model is given below:

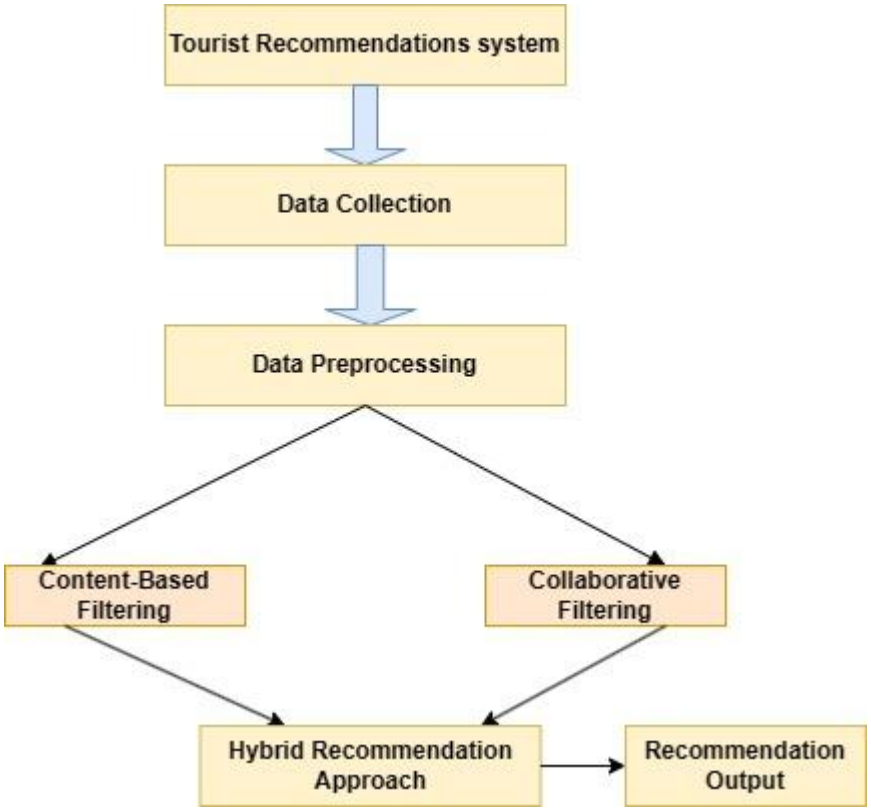

Figure:Proposed model for Tourist Recommendation System

6.3. Algorithm:

Content-Based Filtering:

- Analyse characteristics of places such as points of interest, historical places, and tourist spots.
- Recommend places based on similarity to user preferences, considering factors like category, ratings, and user-defined criteria.

Collaborative Filtering:

- Implement *k* Nearest Neighbors (KNN) algorithm to recommend places by analysing the preferences and behaviours of similar travellers.
- Suggest nearby destinations based on the preferences and behaviours of similar users.

6.4. Evaluation and Optimization:

- Evaluate the performance of the recommendation system using metrics such as accuracy, ratings, and user satisfaction.
- Optimise the recommendation algorithms based on user feedback and interaction data.
- Update the dataset Continuously and refine the recommendation system to adapt to changing user preferences and trends.

By integrating both content-based and collaborative filtering techniques, the proposed recommendation system aims to provide travellers with tailored recommendations that align with their preferences and time constraints, enhancing their exploration experience in Bangladesh's Upazilas.

7. Dataset Collection

When we attempt to collect databases we have faced many difficulties which are given below:

- At first, we tried to collect the dataset by collecting the related papers and from different websites but we didn't get any suitable data for the tourist recommendation system.
- Then we try to make a dataset of our own on Bangladesh’s tourist places in every upazila. We tried to use google map api key for data collection with coding and manual process. Here, we tried to use Wikipedia to code in order to gather data using the Google Maps API but it didn't work. We faced many problems as the api key was not freely available. The system needed credit card information and after some limited use of that api key, they would charge money which was costly. As a result, we were forced to use manual data processing by google place api alternative method and it was a time consuming process. But we have collected the dataset successfully with google place api key.

- Then, when we were storing data in the database where single quotes (') caused a problem for us when we attempted to import the first 20 datasets into PHP MySQL. The sql format does not support (''). Therefore we have applied some methods in the code to remove them datawise. After that the data was imported successfully.
- We have also tried to import a huge set of data in the sql with only code at a time but the limitation was it was not taking the whole dataset in the file, it was unable to read the dataset from the file so we have finally imported the dataset by copy paste method folder wise.
- MySQL was crashing a lot when we attempted to import a large dataset—roughly 15,000 datasets—into it. In order to resolve this problem, we had to change the MySQL configuration.After then the crashing has been stopped.

After facing the problems, we have collected data by remaining steps. The dataset preparation process has required researchers to follow best practices to collect data using the Google Places API, processing it into a suitable format, and storing it in a MySQL database. Here's a detailed explanation how we are collected data each step:

7.1. Data Gathering from Google places API:

- Firstly, we have collected data by utilising the Google Places API to retrieve information about famous places in Bangladesh's Upazilas. We have used specific keywords related to "points of interest," "historical places," and "tourist places." to retrieve data for each upazila based. We got the name of all upazila of Bangladesh in this website: https://www.bangladesh.gov.bd/site/view/upazila-list&lang=en .
- We have stored the obtained information in distinct JSON files, each of which should contain information on around 10 places inside each upazila.

7.2.Data Processing:

- We have handled JSON format data by using python script.
- We have verified that each of the text files in each raw dataset contain the expected JSON format.
- To make sure that the data is completed, fill in any missing values with the appropriate default values.
- We have processed data to collect the necessary information, such as the title (place name), upazila name, keyword, longitude, latitude, rating, rating count, category, ID, and search key, which must be extracted from each JSON file.
- We also remove unnecessary columns which are not needed, including position, thumbnails.

7.3. Data Transformation:

- Once data are evaluated into suitable format for added into a database table,
- Then, we have arranged the data in a tabular format, with the relevant columns representing the extracted features.
- Also, we are ensured that everything is in keeping with the database schema and consistent.

7.4. Database Integration:

- For the purpose to store the processed data, set up a MySQL database.
- We have created database tables with the proper schema to accommodate the extracted features.
- We have used SQL queries to insert the transformed data into the database tables.
- We have handled any possible mistakes or inconsistencies that may arise during the insertion process to maintain data integrity.

7.5. Quality Assurance

- We are checked to make sure all required features are included and formatted correctly.
- We have resolved any inconsistencies identified throughout the validation procedure.

After processing the data, the documentation scenario details are given:

The Datasets directory in the repository comprises eight main folders which are “Chattogram”, “Rajshahi”, “Khulna”, “Barisal”, “Sylhet”, “Dhaka”, “Rangpur” and “Mymensingh”. Each of these folders contains two sub-folders with the following names: “point of interest” and “Tourist place and Historical places”. The “point of interest” sub-folder of Chattogram contains 103 text files and “Tourist place and Historical places” of Khulna contains 206 text files likewise “point of interest" (Rajshahi)- 66 text files, “Tourist place and Historical places”(Rajshahi)- 132 text files, “point of interest" (Khulna)- 59 text files, “Tourist place and Historical places”(Rajshahi)- 118 text files, “point of interest" (Barisal)- 42 text files, “Tourist place and Historical places”(Rajshahi)- 84 text files, “point of interest" (Sylhet)- 41 text files, “Tourist place and Historical places”(Rajshahi)- 82 text files, “point of interest" (Dhaka)- 88 text files, “Tourist place and Historical places”(Dhaka)- 176 text files, “point of interest" (Rangpur)- 58 text files, “Tourist place and Historical places”(Rangpur)- 116 text files, “point of interest" (Mymensingh)- 35 text files, “Tourist place and Historical places”(Mymensingh)- 35 text files. A maximum of ten data entries can be found in each text file.

id	search_key	place_name	title	address	latitude	longitude	category	rating	ratingCount
1	point of interest	Muktagacha	বৃষ্টি সাগর	Q795+G3R	24.7689	90.2576	Tourist attraction	3.5	4
2	point of interest	Muktagacha	মুক্তাগাছা জমিদার বাড়ি, মুক্তাগাছা, ময়মনসিংহ	Q792+2WP, College Rd	24.7693	90.255	Historical landmark	4.3	2380
3	point of interest	Muktagacha	নুর মোহাম্মদ,রেন্ট এ কার	2210 Mymensingh - Tangail Hwy	24.7615	90.2681	Tourist attraction	5	2
4	point of interest	Muktagacha	মেম্বার ফিস ফিড	ডৌয়াখলা, মালতিপুর রোড়	24.766	90.2561	Park	5	1
5	point of interest	Muktagacha	আব্দুল মালেক মঞ্জিল	Q787+WGV	24.7673	90.264	Garden	-9	0
6	point of interest	Muktagacha	Paikashimul Notun Bazar	Q772+6J9	24.763	90.2516	Hiking area	-9	0
7	point of interest	Muktagacha	Talha garden	8 Fulbaria-Muktagachha Rd	24.7617	90.2582	Garden	4	3
8	point of interest	Muktagacha	ধীরে বহে মুক্তাগাছা শিশু পার্ক	Q7C3+9XP	24.771	90.255	Park	3.9	101
9	point of interest	Muktagacha	রাজবাড়ী নৃত্য ঘর	Q793+PR2, কলেজ রোড	24.7693	90.2545	Tourist attraction	5	3
10	point of interest	Muktagacha	কমলাপুর(বড়বাড়ি)	কমলাপুর (বড়বাড়ি	24.7642	90.257	Garden	3	2

Figure: Database with tourism data of every upazila of Bangladesh.

8. Expected Result

Our project will develop a tourist place recommendation system that considers a traveller's specified time frame for travel. This system will recommend tourist destinations that can be visited within that time frame, ensuring effective use of traveller’s available time. Also, this system will plan the optimal route from the traveller's present location to all the recommended destinations, maximising the number

of attractions that can be visited within the given time constraints. So, we will develop a time conscious personalised effective trip planning tourist recommendation system that will make travellers explore the beautiful places within their limited time.

9. Reference:

1. ***Traveler's recommendation system using data mining techniques*. (2018, August 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8697862/references#references**
2. ***Tour Spot Recommendation System via Content-Based Filtering*. (2022, December 14). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10016820**
3. **Sutomo, R., & Pratama, D. (2023). Measuring Tourist Experience in Semarang City through an Advanced Recommendation System. *Jurnal Komunikasi Sains Dan Teknologi*, *2*(2), 192–200. https://doi.org/10.61098/jkst.v2i2.56**
4. **Chougule, V. (2021). Tourist place recommendation System using Machine learning. *ResearchGate*. https://doi.org/10.15680/IJIRSET.2021.1006142**

10. Appendix

CO	CO Descriptions	PO Descriptions	Learning Domains	Assessment Weight	Learning Subdomains	Rubrics Design	Points	Remark
CO1	Integrate new and previously acquired knowledge for identifying a real-life complex computer science and engineering problem as the capstone project.	PO1: Engineering Knowledge	Cognitive	45%	C2	Able to understand the knowledge of mathematics, natural sciences, engineering fundamentals, and computer science and engineering.	20	yes, by collecting dataset, learning related paper and after applying then we know the use of AI in system, system design and use of python and also methodology and use of cse fundamentals.
					C3	Able to apply new and previously acquired knowledge for solving problems.	25	yes, use python coding for data collection, also use methodology to make this recommendation system. Use data science and Ai knowledge.
CO2	Examine various problem domains (literature review), define the problems,	PO4: Investigation	Cognitive, Psychomotor	45% 10%	C4	Able to analyse and/or compare and/or categorise investigation (experiment/modelling/survey/etc.) data.	15	We found by analysis previous related work the limitation and comparison and models that can be used. Also, different data collection techniques that can be

	and formulate the objectives for the capstone project.							**gathered from paper , website or by your own surveys.**
					C5	**Able to critically evaluate and/or assess and/or interpret and/or make inferences and/or conclusions data from the investigation (experiment/modelling/survey/etc.).**	**15**	**Able to know different data collection techniques that can be gathered from paper , website, by your own like surveys,coding etc..**
					C6	**Able to design investigation methodology based on literature review and/or standards.**	**15**	**Yes, we have read many papers on investigation on methodology based on literature review and/or standards.**
					P2	**Able to conduct investigation (experiment/modeling/survey/etc.) under some supervision.**	**3**	**Yes, our supervisor has guided us very well in data collection, what type off project will be good and what methodology we should use.**
					P3	**Able to perform investigation (experiment/modeling/survey/etc.) under minimum supervision.**	**4**	**Yes, we gather data, methodology, and solve problems on our own under minimum supervision.**

					P4	**Able to adapt investigation (experiment/modelling/survey/etc.) to suit objectives.**	**3**	**Yes, we have got a dataset and methodology that suits our objective.**
Total Points							**100**	