Bachelor of Science in Computer Science & Engineering

Development of an Intelligent Tourist Place Recommendation System Using Machine Learning

by

Azmain Abid Khan

ID: 1704130

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

May, 2023

Development of an Intelligent Tourist Place Recommendation System Using Machine Learning

Submitted in partial fulfilment of the requirements for
Degree of Bachelor of Science
in Computer Science & Engineering

by

Azmain Abid Khan

ID: 1704130

Supervised by

Dr. Mahfuzulhoq Chowdhury

Associate Professor

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

The thesis titled '**Development of an Intelligent Tourist Place Recommendation System Using Machine Learning '** submitted by ID: 1704130, Session 2017-2018 has been accepted as satisfactory in fulfilment of the requirement for the degree of Bachelor of Science in Computer Science & Engineering to be awarded by the Chittagong University of Engineering & Technology (CUET).

Board of Examiners

Chairman

Dr. Mahfuzulhoq Chowdhury
Associate Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Member (Ex-Officio)

Dr. Abu Hasnat Mohammad Ashfak Habib
Professor & Head
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Member (External)

Dr. Md. Mokammel Haque
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Declaration of Originality

This is to certify that I am the sole author of this thesis and that neither any part of this thesis nor the whole of the thesis has been submitted for a degree to any other institution.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. I am also aware that if any infringement of anyone's copyright is found, whether intentional or otherwise, I may be subject to legal and disciplinary action determined by Dept. of CSE, CUET.

I hereby assign every rights in the copyright of this thesis work to Dept. of CSE, CUET, who shall be the owner of the copyright of this work and any reproduction or use in any form or by any means whatsoever is prohibited without the consent of Dept. of CSE, CUET.

Signature of the candidate

Date:

Acknowledgements

It is very difficult to complete a task without the grace of Allah Almighty and also the support, supervision, and guidance of several people. First and foremost, I express my gratitude to the Almighty Allah for providing me with the determination and ability to complete my project.

After that, I would like to thank my supervisor, Dr. Mahfuzulhoq Chowdhury, Associate Professor, Dept. of CSE, CUET, from the bottom of my heart for his constant guidance, endless encouragement, cooperation, constructive criticism, helpful suggestions, and constant support during the work of my project. I could not have completed this project on time without his constant and intense support and cooperation.

I would like to express my special gratitude to my friends to provide the necessary information to create the dataset for this project work. Besides Trip Advisor, Gozayaan, and Travelers of Bangladesh Facebook page admin have also helped me a lot in getting my important tour-related data. I also want to express gratitude to our respective parents and family members for the unceasing encouragement, support, and attention throughout my whole life.

Abstract

In the tourism industry building a personalized travel recommendation system is important to enhance the satisfaction and experience of travelers. In this thesis, an automated recommendation system has been developed for a mobile application. It provides two major functionalities. At first, a prediction model has been developed to predict the suitability of a travel destination based on the user's preferences and some other relevant factors. Then the system offers personalized recommendations for the best local places to visit, hotels to stay in, transportation services to use, travel agencies with necessary details, and various activities to do. This system addresses the challenges faced by travelers in planning their trips, like difficulty in making informed decisions, information overload, and lack of personalized recommendations. It utilizes various machine learning classification algorithms to predict the best-suited travel destinations and also provides recommendations to the users, based on their preferences. In this study, real tourism-related data from different sources were collected to create a dataset, and several algorithms were applied to the dataset for destination and activity recommendation systems. The proposed system has the potential to improve the overall tourism experience for travelers and enhance the competitiveness of the tourism industry. Future work may involve integrating social media data and user-generated content to improve the accuracy and personalization of the recommendations and developing a feature for real-time feedback from users to further enhance the system's performance.

Keywords: Automated recommendation system, personalized travel recommendations, user preferences, machine learning classification algorithms, travel destinations, local places to visit, hotels, transportation services, travel agencies, activities, informed decisions, information overload, lack of personalized recommendations, tourism experience, the competitiveness of the tourism industry, social media data, user-generated content, real-time feedback

Table of Contents

Acknowledgements **iii**

Abstract **iv**

List of Figures **viii**

List of Tables **ix**

1 Introduction **1**

1.1 Introduction 1

1.1.1 What is Tourism? 2

1.1.2 Contribution of tourism to GDP 3

1.1.3 Machine Learning 3

1.1.4 Tourist Place Suitability Recommendations 4

1.1.4.1 Random Forest Algorithm 4

1.1.4.2 Decision Tree Algorithm: 4

1.1.4.3 K-Nearest Neighbor (KNN) Algorithm 5

1.1.4.4 Logistic Regression Algorithm 5

1.1.4.5 Confusion Matrix 5

1.2 Design Overview of Machine Learning-based Personalized Travel Recommendation System and Checking Suitability of Best Tourist Place 6

1.3 Difficulties 7

1.4 Applications 9

1.5 Motivation 11

1.6 Contribution of the thesis 12

1.7 Thesis Organization 12

1.8 Conclusion 13

2 Literature Review **14**

2.1 Introduction 14

2.2 Related Literature Review 15

2.3 Conclusion 17

3 Methodology **18**

3.1 Introduction . . . 18
3.2 Overview of Framework and Detailed Explanation . . . 18
3.2.1 Build ML model on tour-related dataset . . . 19
3.2.2 Create Android App . . . 20
3.2.3 Deploy ML Model on Android . . . 22
3.2.4 API Integration with Android Applications . . . 23
3.2.5 Implementation of the Proposed Method . . . 23
3.2.6 System Requirements . . . 23
3.2.7 Implementation Details . . . 24
3.2.7.1 Steps for the best Tourist Place Suitability Prediction Model . . . 24
3.2.7.2 Steps for Android Application . . . 26
3.3 Conclusion . . . 32

4 Results and Discussions **33**
4.1 Introduction . . . 33
4.2 Dataset Description . . . 33
4.3 Impact Analysis . . . 34
4.3.1 Social and Environmental Impact . . . 35
4.4 Evaluation of Framework . . . 36
4.4.1 Data Visualization . . . 36
4.4.2 Best Parameter for K-Nearest Neighbor Algorithm . . . 38
4.4.3 Best Parameter for Random Forest Algorithm . . . 38
4.4.4 Best Parameter for Logistic Regression Algorithm . . . 39
4.4.5 Mobile Application Framework . . . 39
4.5 Evaluation of Performance . . . 42
4.5.1 Performance Evaluation Using Model Accuracy and Confusion Matrix . . . 45
4.5.2 Performance Evaluation Using AUC-ROC Curve . . . 50
4.5.3 Performance Evaluation for App Recommendation . . . 51
4.6 Conclusion . . . 52

5 Conclusion **53**
5.1 Conclusion . . . 53
5.2 Future Work . . . 54

List of Figures

1.1 Overview of the recommendation system in mobile application . . 8
1.2 System Architecture of an Intelligent Recommendation System (a) 9
1.3 System Architecture of an Intelligent Recommendation System (b) 10

3.1 ML Model Framework 19
3.2 Process Flow of Android App 21
3.3 Deploy ML Model on Android 22
3.4 User database 24
3.5 UI design of homepage 27
3.6 UI design of Discover button page 28
3.7 UI design of result page (a) 29
3.8 UI design of result page (b) 30
3.9 User authentication database 31
3.10 User database 31

4.1 Features of dataset 34
4.2 Sample portion of dataset 34
4.3 Comparison of features in order of importance 35
4.4 User Rating 36
4.5 Suitable trip data visualization based on tourist friendly and safety 37
4.6 Suitable trip data visualization based on travel guide and attractions 37
4.7 Suitable trip data visualization based on transport and hotel . . . 37
4.8 Model accuracy comparison against neighbors 38
4.9 Logistic Regression and Random Forest Algorithm 39
4.10 Login and Sign up page 40
4.11 Logo and profile section 41
4.12 Homepage section 42
4.13 Gathering information from travelers for checking place suitability 43
4.14 Place suitability is displayed in the mobile app 44
4.15 Firebase Real-Time Database 45
4.16 Recommendations of local places in the mobile app (a) 46
4.17 Recommendations of local places and others in the mobile app . . 47
4.18 Performance Comparison 48
4.19 Confusion matrix of K-Nearest Neighbor Algorithm 49
4.20 Confusion matrix of Logistic Regression Algorithm 49

4.21 Confusion matrix of Random Forest Algorithm 49
4.22 Confusion matrix of Naive Bayes Algorithm 49
4.23 ROC-AUC curve of five algorithms model 51

List of Tables

4.1 Confusion Matrix 43
4.2 The accuracy of several algorithms 45
4.3 The performance of several algorithms 46

Chapter 1

Introduction

1.1 Introduction

Tourism is a crucial contributor to the economic growth and development of many countries, and Bangladesh is no exception. The country is endowed with rich natural beauty and historical sites, such as archaeological sites, mountains, beaches, rivers, hills, forests, waterfalls, tea gardens, and religious places, which attract a large number of visitors annually. The tourism industry in Bangladesh has a significant impact on the country's Gross Domestic Product (GDP), with a contribution of 4.4 percent in 2018-2019 and 4.2 percent in 2016-2017, according to the Bangladesh Tourism Board [1]. This industry has created many job opportunities, increased economic revenue, and improved infrastructure in the country.

Bangladesh is a developing country. Here tourism can bring many environmental, economic, and social benefits. It can earn a large amount of money from international tourists. There are many opportunities to boost the tourism industry in Bangladesh[2]. Here people are interested in traveling and exploring new places. But they have no clear concept of the travel destination, time, duration, etc. Sometimes they have no idea that there are some places nearby. So they can't get the scope of traveling of these places. Budget is also a factor in traveling. People may like budget-friendly or luxury travel [3]. So budget-friendly travelers hesitate to travel to long-distance places, thinking it may cross their budget. Besides, all destinations are not for family or friends. Some are best for camping and sightseeing. Some are familiar with historical values and natural beauty. So a person's personal choice becomes a factor for traveling a place [4].

Various trip management apps and websites such as Trip Adviser [5], Travel

Buddy, and Gozayaan [6] have many facilities in Bangladesh. People can use this facility to travel to various places. But here, suggesting the best destination according to budget, time, and interests is absent. So travelers can't get any suggestions about a destination according to their budget and time. They also don't know what activities they can do there and what foods and restaurants they will try.

Solving these problems at a time is a big challenge. That's why an Android-based smartphone application is being proposed here. It will help travelers to introduce the best tourist place recommendation system in Bangladesh. This application will provide the opportunity to reduce the wastage of time in selecting a destination under budget and time.

In the current world, technological advancements have revolutionized almost every sector, and mobile applications have become indispensable for our daily needs. The proposed tourist recommendation system aims to leverage technology to make travel easier, more efficient, and more personalized. The system's primary goal is to help travelers by providing them with personalized recommendations based on their interests, preferences, and budgets. With this application, potential tourists can explore new places in Bangladesh, experience local culture, and make informed decisions that lead to enjoyable and memorable travel experiences.

1.1.1 What is Tourism?

Tourism is the activity of visiting places for relaxation, happiness, or work purposes. It involves staying in hotels, enjoying attractions, exploring different lifestyles, and doing activities like shopping, hiking, sightseeing, and swimming. Tourism is a modern industry that encircles many sectors like entertainment, the food-processing industry, transportation, and hospitality [7]. In many countries, it generates a remarkable amount of revenue and business and also promotes cultural interchange and international agreements.

1.1.2 Contribution of tourism to GDP

According to the World Travel and Tourism Council (WTTC), the direct contribution of tourism to Bangladesh's GDP was BDT 266.2 billion in 2019, which is approximately 3.3 percent of the country's total GDP [3]. The direct contribution includes the economic activity generated by industries that primarily deal with tourists, such as hotels, restaurants, transportation, and leisure activities.

In addition to the direct contribution, tourism also has indirect and induced contributions to the country's GDP. Indirect contributions refer to the economic activity generated by the supply chain of tourism, such as the purchases of goods and services by the tourism industry from other sectors. Induced contributions refer to the economic activity generated by the spending of tourism industry employees and suppliers.

Overall, the total contribution of tourism to Bangladesh's GDP (direct, indirect, and induced) was BDT 861.4 billion, which is approximately 4.4 percent of the country's total GDP in 2019. Despite its potential, the tourism industry in Bangladesh is still largely untapped, and there is room for growth and development in this sector.

1.1.3 Machine Learning

Machine learning (ML) is a fascinating field that enables computers to handle data more efficiently. Sometimes, despite our best efforts, we may find it difficult to evaluate or extrapolate information from data. In such cases, machine learning can prove to be extremely useful. With the availability of vast datasets, there has been a growing demand for machine learning. It is a field of study that focuses on empowering computer programs to automatically learn from experience and improve their performance. By gathering diverse observations and projected data, machine learning trains computers to carry out various tasks such as predictions, recommendations, and guesses, based on prior experiences or historical data. It is an area of research dedicated to comprehending and developing "learning" processes that use data to enhance performance on given tasks. Machine learning is an integral part of the rapidly expanding discipline of data science.

Statistical techniques are used to train algorithms to produce classifications, and predictions, and to uncover crucial insights in data mining projects. Ultimately, the decisions made as a result of these insights have a significant impact on key growth indicators in applications.

1.1.4 Tourist Place Suitability Recommendations

We can determine if a place will be suitable or not using the prediction of a modernized tourist place recommendation system based on a machine learning technique. Using a classification method would be highly beneficial if the prediction was accurate. For datasets with known class labels, supervised machine learning techniques can be used; for datasets with unknown class labels, unsupervised learning approaches can be. In our work, we have used several machine learning classification algorithms to predict whether a place will be suitable or not for travelers in Bangladesh.

1.1.4.1 Random Forest Algorithm

Random Forest is a popular supervised machine learning technique that can solve both classification and regression problems. It is known for generating multiple decision trees during training time, which then use majority voting to provide more accurate and consistent predictions. This technique creates a classification system that produces a large number of decision trees, each with higher predictive accuracy than the others. Random Forest is an ensemble learning technique that is widely used for classification, regression, and other tasks. In our project, we have extensively tested Random Forest with various parameters to ensure optimal results.

1.1.4.2 Decision Tree Algorithm:

There are various problems in both classification and regression that require effective solutions. To address these issues, one popular technique is to utilize a decision tree algorithm, which falls under the category of supervised learning. The algorithm works by utilizing three types of nodes, namely leaf, internal, and branches. The branches represent the decision rules, internal nodes signify the

features, and leaf nodes represent the final outcome. After training the model, we test it with a variety of data to ensure its effectiveness in real-world scenarios.

1.1.4.3 K-Nearest Neighbor (KNN) Algorithm

The K-Nearest Neighbor (KNN) algorithm is a widely-used supervised classifier in the field of machine learning. It excels in solving classification tasks by leveraging the power of similarity to classify new data points. To do this, the algorithm chooses a value for k, which represents the number of nearest neighbors to consider. KNN can solve both classification and regression problems, and it is considered a "lazy learning" method because it simply stores the data members in efficient data structures like hash tables. The KNN algorithm works by selecting k neighbors, calculating the Euclidean distance between them, sorting them in ascending order, and finally counting the number of data points in each category. In our work, we carefully selected the best k value for the KNN algorithm to achieve optimal performance.

1.1.4.4 Logistic Regression Algorithm

Logistic Regression is a powerful machine-learning technique for modeling data with a binary dependent variable. This method is commonly used to describe the relationship between dependent and independent variables. In cases where the output variable is assumed to be continuous, a simpler linear regression model may suffice. However, when the output variable is dichotomous or not continuous, logistic regression provides a superior alternative. One of the major advantages of this approach is its ability to easily regularize the model, leading to improved accuracy and performance. Additionally, logistic regression is computationally efficient and simple to implement. Unlike some other machine learning methods, scaling input features is not necessary, making it an attractive choice for many applications.

1.1.4.5 Confusion Matrix

The confusion matrix is a valuable tool for assessing the performance of a classification algorithm. When dealing with datasets containing multiple classes or

imbalanced observations, relying solely on classification accuracy can be misleading. In our evaluation of the effectiveness of our classification models on a particular set of test data, we utilized the confusion matrix. The matrix presents four possible combinations of anticipated and actual values, providing a more comprehensive understanding of the algorithm's performance.

1.2 Design Overview of Machine Learning-based Personalized Travel Recommendation System and Checking Suitability of Best Tourist Place

The objectives of this part are divided into two parts:

- Best tourist place suitability recommendation system using machine learning algorithms.
- A mobile application for determining the best local places based on personal preferences.

The overall process is shown in detail in diagram Fig 1.1. This figure represents the system architecture of the Best tourist place recommendation system. The primary objective of the system model is to provide travelers with accurate recommendations on whether a particular place is suitable for them or not. This framework comprises the following main steps:

Data Preprocessing: Preprocessing is required for this entrepreneur dataset. In the stage of data preprocessing, transferring categorical data into a machine-understandable format. We also removed all kinds of values that can cause errors such as incomplete values, redundant values, missing data, etc. Then we have to extract the features.

Create and choose the optimal ML model: To ensure that our model is able to generalize well to new, unseen data, we split our dataset into two subsets: one for training and one for testing. By training the model on the training data

and evaluating its performance on the testing data, we can get a sense of how well the model is able to make accurate predictions on new, unseen data. A tiny portion of the data is used for testing, while the majority is used for training. We have used machine learning algorithms like the Logistic Regression Algorithm, KNearest Neighbor (KNN) Algorithm, Random Forest Algorithm, and Decision Tree Classifier Algorithm to build our models. After that, we have training data to train and we also have tested data to test the model. We have chosen the best model based on accuracy after these models have been trained with training data. Finally, we can predict whether a place is recommended or not for a traveler.

Flask API: The form will be filled out by the user, who will then submit it. The form will then receive a POST request. Additionally, when a user submits a post request, the Flask API will accept their data and send it to the machine learning model so that it can predict the output class. We will send the predicted class as JSON to the Android app.

Android UI with Java: We initially required an Android studio in order to work with Android programming. We are aware that the UI is constantly produced in an XML file. After designing our UI, we wrote backend logic to accept data from the front end in a Java file. Then we connected the API to the Android app. The logic we must use is that after entering data into the Android app and using the API, the Android app will display the API's answer. One issue was that the Android app could not recognize that the API we created was running locally on our system. Therefore, we have deployed our API online, and we would do it using Railway.

1.3 Difficulties

Creating a useful machine-learning model can be challenging. The main challenges are finding a suitable dataset for training the model and selecting the best model for the task. These challenges can be further divided into several objectives.

Data Collection for Creating Dataset: There was no existing tour-related

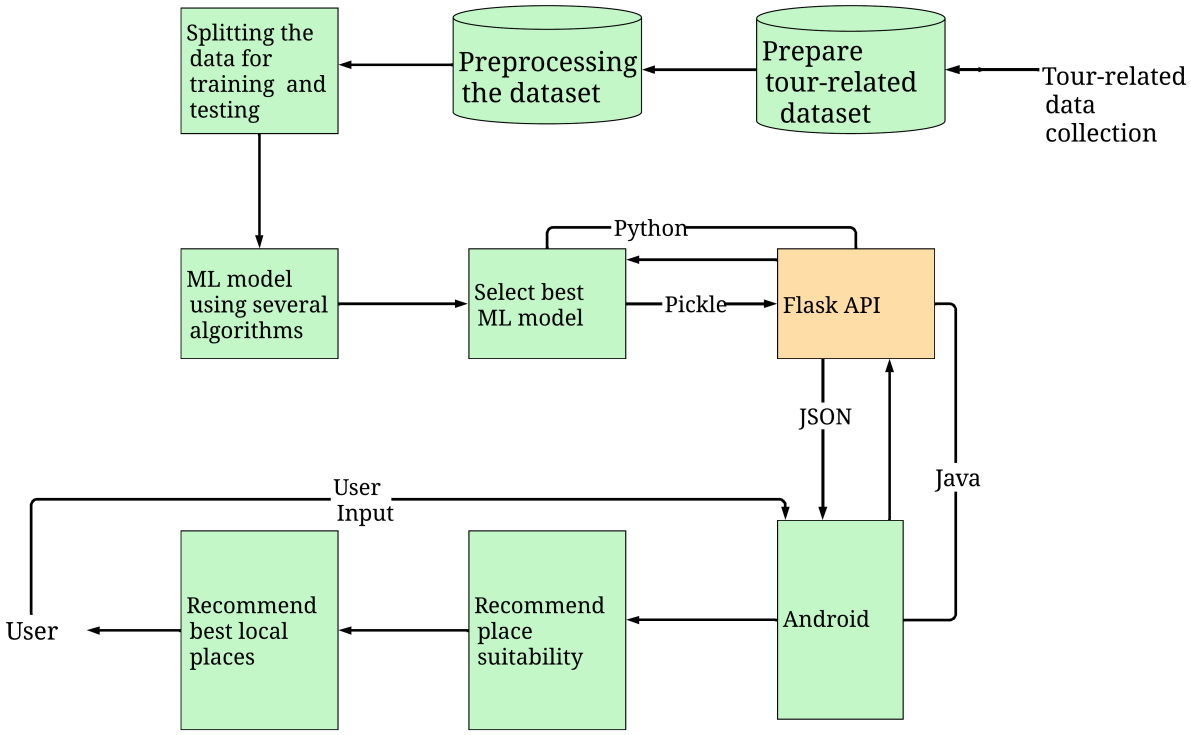


Figure 1.1: Overview of the recommendation system in mobile application

dataset for travelers in Bangladesh. So it was very difficult to collect tour-related data from travelers in Bangladesh for the purpose of creating a dataset. Unfortunately, the tour information a traveler was not enough. So it was very difficult to collect exact information. So I have tried to create the dataset. Fortunately, various travel agencies helped me to create the dataset.

Deploying Machine Learning Model on Android: The basic process for deploying machine learning on a website entails building the model in any Python IDE, extracting it using a pickle module, and then deploying it as a web app with the use of any web framework, such as flask or. As our machine learning model's output will be in JSON format, we have developed a Flask API. We have deployed machine learning models on a website, Railway. This implementation process was slightly difficult for me.

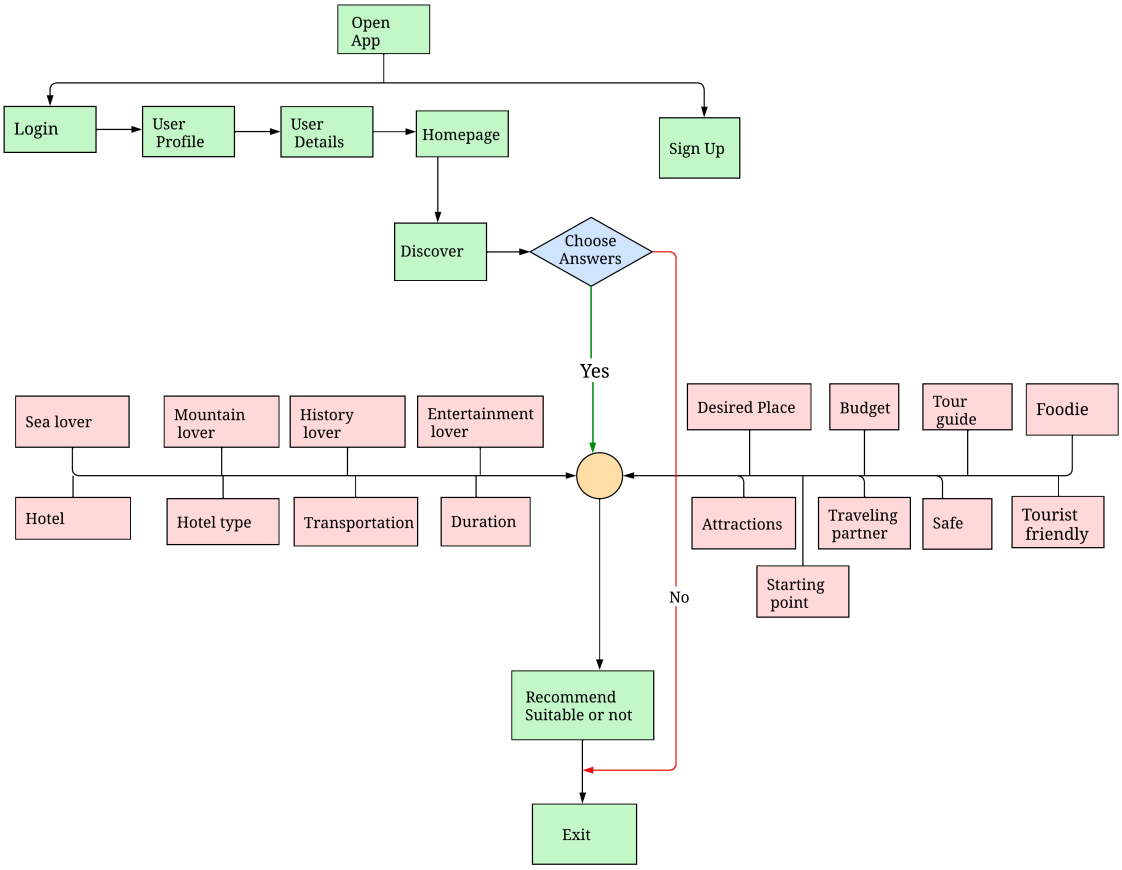


Figure 1.2: System Architecture of an Intelligent Recommendation System (a)

A Recommendation for a Suitable Place for Travelers Based on User Data: Generating recommendations automatically is a difficult issue. It is difficult to recommend a suitable place for travelers as per their expectations, location, interests, etc. To recommend a suitable place for travelers in Bangladesh, I have compiled a list of options that are required to know about the personal choices and interests of a traveler.

1.4 Applications

The system model has been created to predict whether a place is suitable or not for a traveler. Therefore, in my opinion, the suggested system can be used in other dimensions. The following are a few examples of potential applications:

- This system can help travelers plan their trips before traveling based on

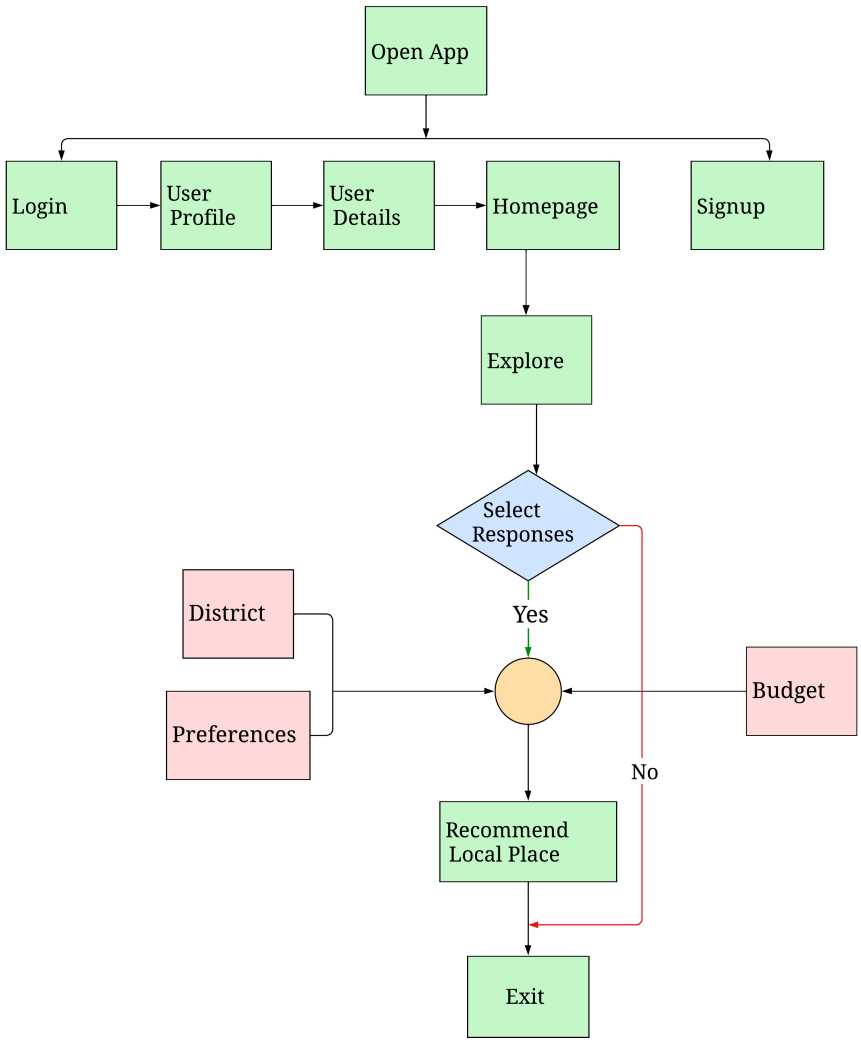


Figure 1.3: System Architecture of an Intelligent Recommendation System (b)

their budgets and preferences. It can also help travelers make correct decisions about their plans by providing personalized recommendations for attractions, destinations, and accommodations.

- To promote tourism in Bangladesh this system can be helpful. It can attract travelers around the world by advertising the best attractions and destinations. Thus it will help to boost the local economy here.
- New business opportunities will be generated by this system to encourage unfamiliar attractions. Thus it can boost business in the tourism industry [8].

- Cultural exchange between local communities and travelers can be facilitated by this system. Travelers will learn about various traditions and cultures using this system which will recommend destinations of different cultures and traditions.

1.5 Motivation

Tourism is an important sector in many countries and it is significantly related to economic development[9]. Nowadays mobile applications have become an important tool of tourism with technological advancements. Travelers can experience a unique experience and enjoy their trip using a personalized travel recommendation system. But building an effective and efficient personalized travel recommendation system is not easy. It has many challenges like collecting the appropriate dataset and selecting the best model. For addressing these challenges here we propose a personalized travel recommendation system for a mobile application. Our proposed system aims to provide personalized recommendations to travelers based on their preferences and interests. Thus it will make their travel experience more enjoyable and personalized. Here we will evaluate the system's performance and its contribution to the field of personalized travel recommendation systems will also be discussed. The tourism industry is growing rapidly and thus it has led to an increase in demand for personalized travel recommendations. However, finding the right recommendations can be a daunting task, especially for travelers who are unfamiliar with the destination they are visiting. To address this issue, we propose a personalized travel recommendation system in the form of a mobile app that utilizes machine learning algorithms to suggest customized travel itineraries based on user's preferences and past behavior. This system has the potential to revolutionize the way people plan and experience their travels, making it more efficient, enjoyable, and tailored to their unique needs. Our work contributes to the field of travel recommendation systems by providing a novel solution that takes advantage of the latest advancements in machine learning techniques.

1.6 Contribution of the thesis

Every thesis or piece of study has been done in order to identify a distinctive contribution. It's not anything like integration or information collecting; rather, it's something like invention, highlighting a cause and effect, or suggesting a model that would benefit people. The thesis works' primary contributions are as follows:

- To gather the information that will be useful for future work.
- To collect tour-related data for creating a dataset.
- To select the ideal machine learning predictive model for the dataset.
- To develop a mobile application based on machine learning that can predict whether a place is suitable or not.
- To develop an automated system that can recommend the best local places based on personal preferences.

1.7 Thesis Organization

The following is the organization of the rest of the thesis:

Chapter – 2

It provides a summary of the previous works in the related fields and its background.

Chapter – 3

This section presents a detailed description of the proposed system and provides a comprehensive overview of its functionality, offering a clear understanding of how the system works.

Chapter – 4

It shows all about the evaluation and experimental outcomes of the proposed system.

Chapter – 5

In addition, a brief outline of future work has been provided, which makes the proposed system more efficiently.

1.8 Conclusion

This chapter provides an introduction to our work on a personalized travel recommendation system in a mobile app. It includes an overview of the proposed framework, the motivation behind the thesis, and its potential applications. The challenges we faced during the development process and the contributions of our work are also discussed. Additionally, this section provides details about the system's performance and functionality.

Chapter 2

Literature Review

2.1 Introduction

Tourism has a crucial impact on the economic growth of a nation. It is a rapidly growing industry worldwide [10]. Travelers are becoming more enlightened and discerning in selecting choices with the increasing availability of technology and information. Planning a trip is much more complex because many factors have to be considered like budget, attractions, accommodations, destination, and personal choices and interests [11]. So, there is a growing need for a decision support system for helping tourists by recommending personalized travel destinations. The tourist place suitability recommendation model is one such decision support system that can provide personalized travel recommendations for travelers based on their choices and requirements. Thus time and effort can be saved by this model while improving the overall travel experience. In a tourist place suitability recommendation model, the criteria can vary depending on the purpose and specific context of the model. In this literature review, we will mainly focus on models that use various criteria such as preferences for mountain or sea environments, desired duration, interest in entertainment or history, budget constraints, accommodation needs, safety concerns, and desire for a tour guide. We will review the existing literature on tourist place suitability recommendations models including their underlying techniques and methodologies, the opportunities and challenges for future research, and the effectiveness of different criteria in predicting the best tourist preferences. Ultimately, this literature review aims to give a comprehensive overview of tourist place suitability recommendation models.

2.2 Related Literature Review

In the work of [12], A Smart Tourist Guidance System has been developed in Bangladesh using Machine Learning. Here Google Maps API is used it finds out the location of all points of interest. Then this model finds the best routes one by one to reach the destinations. But this model depends on the user's GPS location and the accuracy of this model is not shown. So we can't understand how good the model is. This proposed system will suggest the suitability of the best tourist place and predict the local places based on per person budget and other features. Besides, the accuracy will be shown to make the best model.

In the work of [13], A machine learning and deep learning-based Tourist Place Recommendation and Recognition System has been developed. Here user's interest is taken as input by the mobile app and then recommends attractions, restaurants, and hotels using K-mode clustering. It also helps to recognize new places using a Convolutional Neural Network. But the accuracy of this model is below 50 percent. This proposed system will use various machine learning algorithms to increase the accuracy.

In the work of [14], A Machine Learning Based personalized hybrid tourism recommender system has been developed. Here first they collect users' preferences, profiles, and appreciation of previously visited places. Then they recommend to the user the best attractions in a particular place. This proposed system will count on starting point, budget, and duration of time to suggest the suitability of the best tourist place.

In the work of [15], A Tourist Place Recommendation System has been developed using Machine Learning. Here the user will be recommended the new places based on the places he reviews on the system. So this model will not be useful when tourists want to know about the budget. This proposed system will take per person budget as the main feature.

In the work of [16], A Recommendation System Based on Tourist Attraction has been developed. They first collect the visiting history of the user's neighbors. Then the recommendations for attractions are constructed using the Cosine

method. But the model is not shown any accuracy and there is no Android app to use it. This proposed system will use an Android app to make it user-friendly. Machine learning algorithms will also be used here to make this model supervised.

In the work of [17], a machine learning model has been used to predict tourists' responses to a tourist attraction. It records responses from tourists to understand positive or negative judgments about a certain attraction. The main limitation of this paper is they didn't show the accuracy of this model. So we can't understand how good the model was. This proposed system will show which algorithm has the best accuracy.

In the work of [18] they discussed a personalized cruise travel recommendation system. It uses data mining and the authors point out the limitations of the single GPS. They optimize the designed model based on the recommendation patterns. In our proposed system we have used machine-learning algorithms and it predicts suitability and recommends the best places.

In the work of [19] they designed a hybrid collaborative filtering travel recommender system that offers personalized tourist spots based on the user's ratings and preferences. Here they used data from only one travel agency. But in our work, we have gathered data from various travel agencies to clarify the difference in user preferences.

In the work of [20] a travel recommendation system based system has been developed based on social media activity, specifically analyzing users' Twitter data. This system uses a machine learning classifier to identify travel-related tweets and incorporates time-sensitive recency weight to provide personalized recommendations based on the user's most recent interests. The proposed model has shown overall accuracy of 75.23%. But In our proposed system we have used the random forest algorithm to get better accuracy. Besides here various algorithms have been tested and a comparison has been shown such as confusion matrix and Roc Auc curve to get better ideas of these algorithms.

2.3 Conclusion

A literature review offers a detailed analysis and summary of existing research and knowledge on a specific topic. In the tourist place suitability recommendation model, the literature review should focus on various criteria to evaluate the suitability of travel destinations like preferences for mountain or sea environments, desired duration, interest in entertainment or history, budget constraints, accommodation needs, safety concerns, and desire for a tour guide, etc. This model can help travelers to improve their overall travel experience and save effort and time. However, like any system, there may be some challenges and limitations associated with the implementation of a tourist place suitability recommendation model. It may include issues related to the accuracy of the model, updates, and maintenance of the system, and data availability. Overall, a literature review can help to figure out the challenges and opportunities associated with the implementation of a tourist place suitability recommendation model. By using the newest technology and research, it is possible to build a system that can benefit both the tourism industry and tourists as a whole.

Chapter 3

Methodology

3.1 Introduction

The systematic approach to solving a research topic through the collection of data using various approaches, the provision of an interpretation of the data collected, and the drawing of inferences from the study data are known as a methodology in research. A research technique is essentially the design of a research or study. It all depends on a researcher's methodical approach to obtaining accurate and trustworthy results in order to guarantee the acceptability of the work. I have undergone a number of processes for the suggested model. Data has already been gathered from a number of travel agencies. Some studies have been conducted for the purpose of creating questionnaires. Finally, the development, testing, and checking of place suitability for travelers in Bangladesh using machine learning.

3.2 Overview of Framework and Detailed Explanation

The suggested system for best tourist place suitability recommendation and determining the best local places based on personal preferences will work on mainly three steps:

1. Build an ML model on the tour-related dataset
2. Create Android App
3. Deploy ML Model on Android

3.2.1 Build ML model on tour-related dataset

For recommending the suitability of tourist places, we have built an ML model using several classification algorithms.

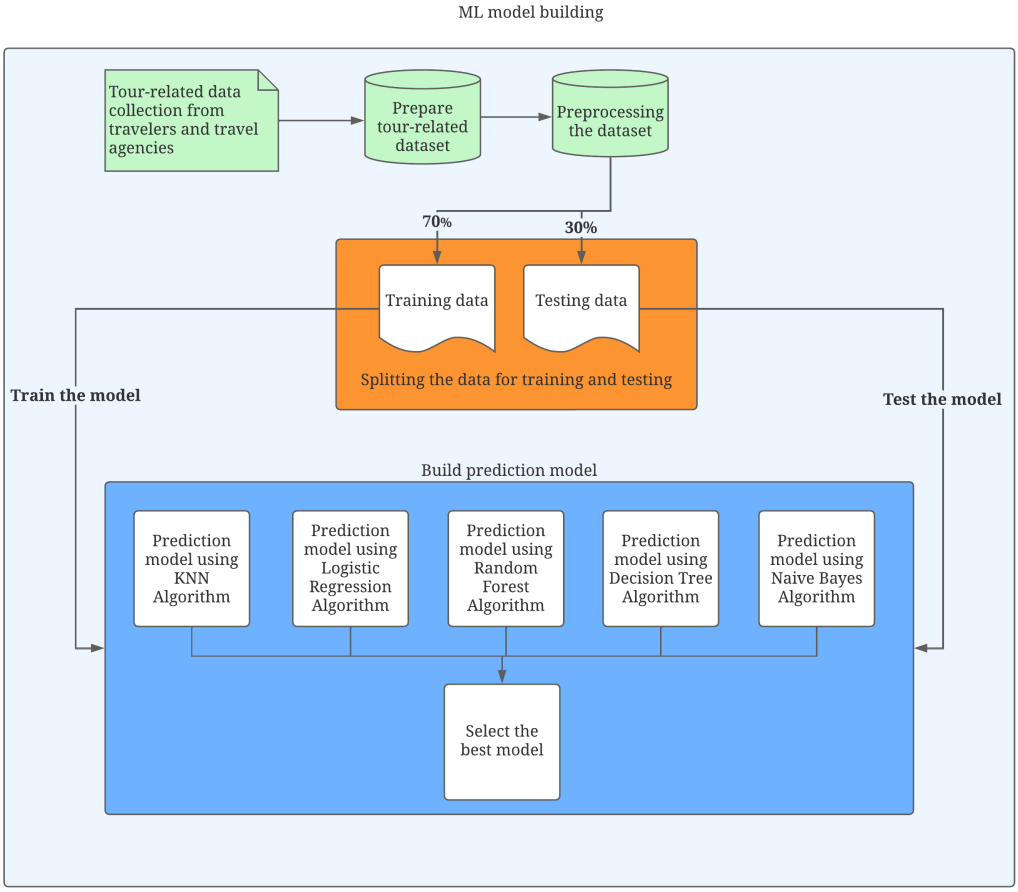


Figure 3.1: ML Model Framework

Figure 3.1 represents the ML model framework. First, we collected data from travelers and travel agencies and then created a dataset. The features of this dataset were: Sea lover, mountain lover, history lover, entertainment lover, need hotel, hotel type, need transport, days, place, budget, travel guide, prefer attractions, traveling partner, prefer safety, foodie, tourist-friendly place, starting point. The data was split into training and testing sets, with 70% of the data used for training and 30% for testing. We have used machine learning algorithms like the Logistic Regression Algorithm, KNearest Neighbor (KNN) Algorithm, Random Forest Algorithm, Decision Tree Classifier Algorithm, and Naive Bayes to build our models. After that, we have training data to train and we also have

testing data to test the model. We have chosen the best model to recommend the suitability of tourist places based on accuracy after these models have been trained with training data and tested with testing data. In our work, the model using the Random Forest algorithm was the best model. Finally, we can recommend the suitability of tourist places using this model. Besides, it is also used to recommend other factors.

3.2.2 Create Android App

We initially required an Android studio to work with Android programming. We are aware that the UI (User Interface) is constantly produced in an XML file. After designing our UI, we wrote backend logic to accept data from the front end in a Java file. Our Android app name is TourMate. Firebase serves as a backend system that connects mobile apps and web applications to cloud storage and APIs. Figure 3.2 represents the process flow diagram of our app.

First, if any user wants to use the system, he/she has to open the TourMate app, and a splash screen will be displayed. If the user has already logged in to this app, then he/she can enter directly into the home fragment. Otherwise, he/she has to enter the login fragment. To log in to this app, an email and password are required. And if the email is valid, then the home activity interface is displayed. If the traveler is new, he has to sign up first. To sign up for this app, a username, email, mobile number, occupation, password, and confirm password are required. If the password is matched with the confirmed password, then the home activity interface is displayed. In-home activity, three interfaces such as Explore, Discover, and User Profile will be found. Travelers can see their personnel information in the User Profile interface, and they can change their information if required. They can also log out of the system if they want. If an individual traveler wants to check the suitability of a place, he/she needs to enter the Discover interface and have to fill in the required inputs such as Sea lover, mountain lover, history lover, entertainment lover, or not, need a hotel or not, hotel type, need transport or not, days, desired destination, budget, need travel guide or not, prefer attractions or not, traveling partner information, prefer safety or not, foodie, prefer a tourist-friendly place or not, starting point. Here,

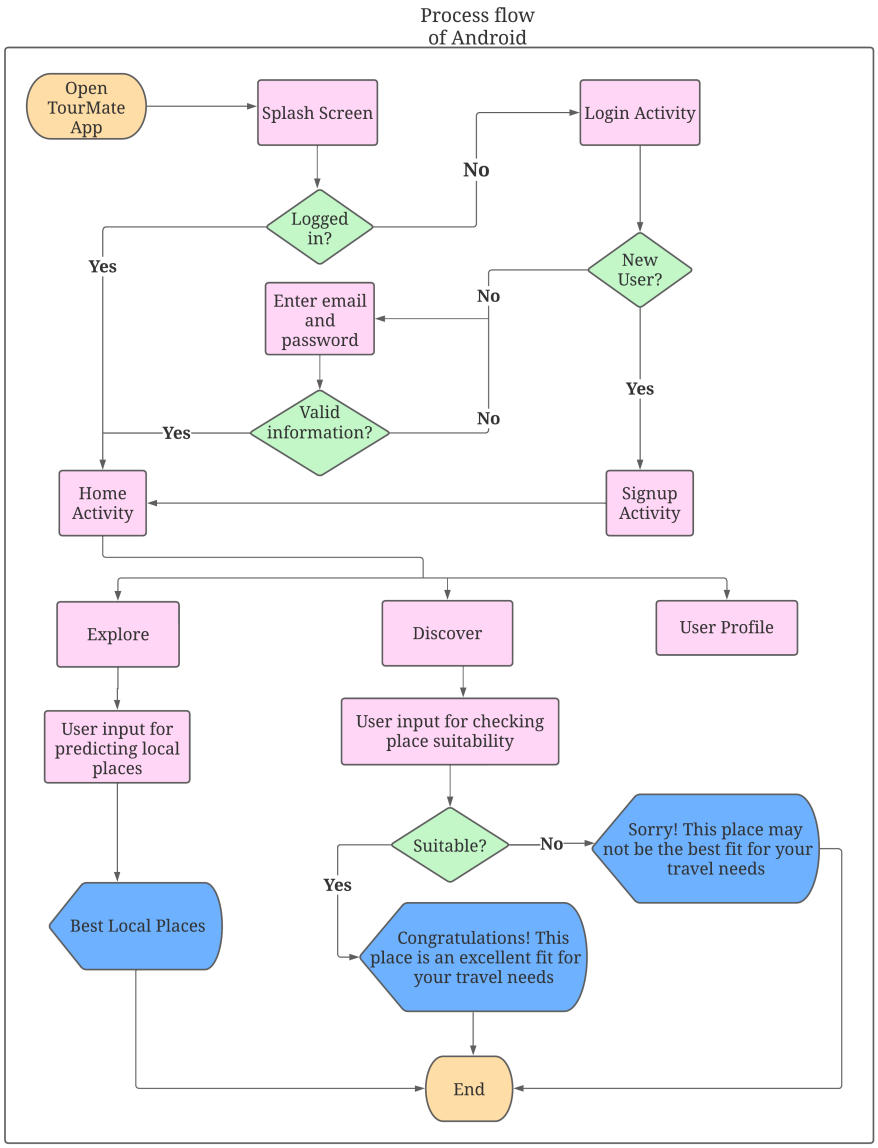


Figure 3.2: Process Flow of Android App

the suitability of the destination is checked using a machine-learning model. If one place is perfect, it will be recommended according to one's demand. If it is not suitable for a traveler, no place will be recommended as suitable. After the result, the users can get back to the home page and can also check other factors. In Explore interface, they will see some questions such as budget, user choice, and desired destination(district). By answering them, this system will recommend to them the best local places there with the necessary information.

3.2.3 Deploy ML Model on Android

First, we saved our machine-learning model as an a.pkl file. Then we have the flask API using Python in Pycharm. Figure 3.3 represents the deployment of the ML model on Android.

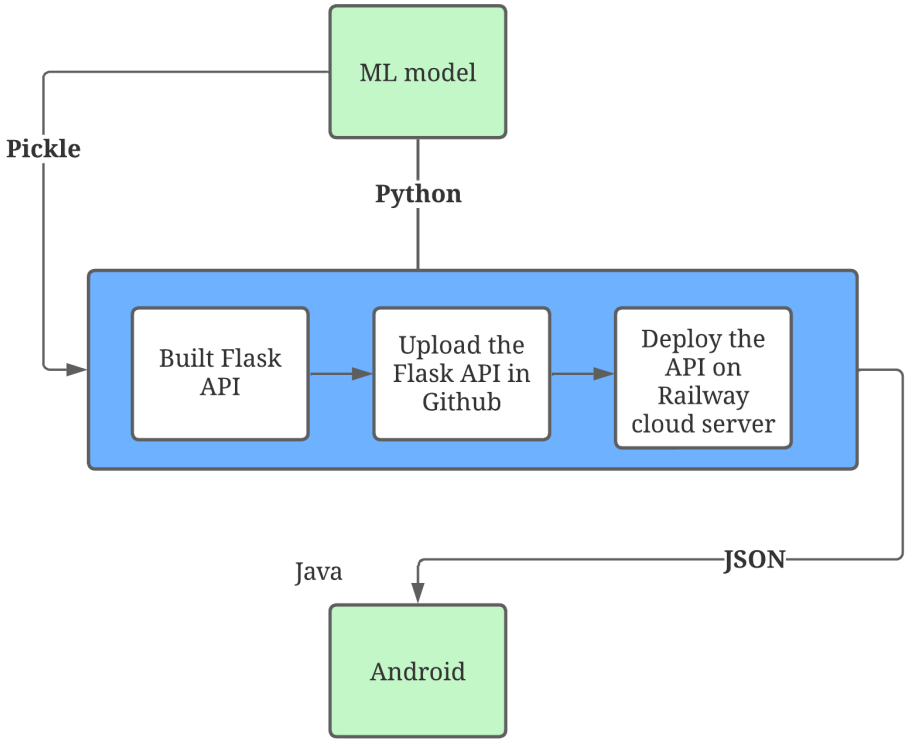


Figure 3.3: Deploy ML Model on Android

The form will be filled out by the user, who will then submit it. The form will then receive a POST request. Additionally, when a user submits a post request, the Flask API will accept their data and send it to the machine learning model so that it can predict the output class. We will send the predicted class as JSON to the Android app. We have used Postman to verify the API of our TourMate App. An automated and interactive tool called Postman is used to check the project's APIs. One library called Volley is required to hit API.

3.2.4 API Integration with Android Applications

The logic we have created is that we will use the Android app's inputs to call the API, and then the API's answer will be displayed back in the Android app. So, there was one issue that the Android app was unable to detect because the API we developed was running locally on our system. Therefore, we had to launch our API online, and Railway was used to do so.

3.2.5 Implementation of the Proposed Method

Combining all the thoughts and models is the last phase. We have completed the setup in this section. The majority of the work in this area has been finished with the system design implementation section. We also provided a thorough explanation of the data collection process and the creation of Android applications, two of the most crucial aspects of our project. This portion will offer a sequential direction while the other process accumulates. There are two parts to the entire implementation procedure. The proposed system involves two steps: developing a prediction model to suggest whether a place is suitable for a traveler and providing recommendations for local places based on user data and the prediction model.

3.2.6 System Requirements

Good hardware is a requirement for completing any computing task effectively and in a reasonable amount of time. A proper programming language is required to carry out any project on a computer. Here is a list of the equipment needed to put our system into practice:

- Operating System: Windows 11 with 8 GB RAM,
- Google Colab for machine learning code visualization,
- Android Studio for App development,
- PyCharm IDE for API Flask development,
- Firebase Realtime Database used for storing data and also for authentication,
- Python is used for model prediction and API development,
- The XML language is used for the front-end design,

- The Java language is used for controlling the front end,
- An Android phone to run our Android project.

3.2.7 Implementation Details

The overall system implementation part is divided into several parts. We will discuss all these parts in the below subsection.

3.2.7.1 Steps for the best Tourist Place Suitability Prediction Model

This subsection will briefly cover the method of developing the best Tourist Place Suitability prediction model. Figure 3.1 represents the overall steps of the prediction model.

Step 1: Firstly, we have collected information on tour-related data from several travelers and travel agencies. I have tried to gather the dataset since the very beginning. After collecting tour-related data from travelers and travel agencies, I created a tour-related dataset. Figure 3.4 represents the tail data of the tour-related dataset.

	D	E	F	G	H	I
1	Is that place right for a m	Is that place right for a hi	Available Entertainment F	Hotel needed	Transport needed	Days
16	1	0	1	0	1	1
17	1	0	1	1	1	1
18	1	0	1	1	1	2
19	1	0	1	1	1	2
20	1	0	1	1	1	2
21	1	0	1	1	1	2
22	1	0	1	1	1	2
23	1	0	1	1	1	1
24	1	0	1	1	1	2
25	1	0	1	1	1	1
26	1	0	1	1	1	2
27	1	0	1	1	1	2
28	1	0	1	1	1	2

Figure 3.4: User database

As this information is a very confidential issues for travel agencies, which is why I cannot collect a large number of data points. 868 instances and 17 features are present in this dataset.

Step 2: The figure shows that in this dataset on travelers, there were some categorical information. The categorical data cannot be understood by a machine.

In order to create numerical data, we convert categorical data using label encoding. Then we determine whether a value is missing or not. Fortunately, our dataset did not contain any missing values. Therefore, handling the missing value wasn't necessary in our dataset.

Step 3: Feature selection is a technique that involves reducing the input variables for a model by selecting only relevant data while eliminating irrelevant data, and there are numerous features that are relevant to predicting the suitability of tourist destinations. To determine the crucial features in our dataset, we used the feature importance technique. We have imported ExtraTreesClassifier, an ensemble method, in order to apply this function.

Step 4:

In the model training phase, the dataset is split into 70% for training and 30% for testing, and various machine learning classification algorithms such as logistic regression, k-nearest neighbor, random forest, and decision tree classifier are used to train the model, and their accuracy is evaluated to choose the most effective classifier for the dataset of travelers.

Step 5: We have selected the best parameters for these algorithms.
1. For the Random Forest Algorithm, its default parameters were:

```
(bootstrap = True,
max_features = 'auto',
min_samples_leaf = 1,
n_estimators = 100)
```

The accuracy of the result was very low at this default parameter. After hyperparameter tuning, we have changed some parameters, such as:

```
  (bootstrap=False,
max_features='sqrt',
min_samples_leaf=2,
n_estimators=30).
```

2. For the K-Nearest Neighbor Algorithm, its default parameters were:

```
    (n_neighbors=5,
weights='uniform',
algorithm='auto',
leaf_size=30, p=2,
metric='minkowski',
metric_params=None).
```

The accuracy of the result was very low at this default parameter. Then we want to try to change the default number of neighbors. I tested the accuracy between one and thirty neighbors, and the accuracy was highest when the number of neighbors was eleven. The changed parameter was n neighbors = 4.

3. The Logistic Regression Algorithm's best C has been examined. Because we are more interested in getting higher sensitivity, we employ Specifiedscore as recall.

Step 6: It is the last stage of choosing a model. The most effective model has been chosen based on the evaluation matrices. After training the classification algorithms, we used the grid search method to fine-tune their hyperparameters and optimize their performance for the dataset of travelers. In our model, the random forest algorithm gives us the best result compared with the other algorithms.

3.2.7.2 Steps for Android Application

After selecting the best machine learning model for the best tourist pace suitability prediction, we have predicted the suitability of the place, and the result is shown in the mobile application. Then we also recommend the best local places, hotels, transportation services, travel agencies, and various activities to do based on personal preferences according to their demands via mobile application. The figure represents the process flow diagram of our app.

Step 1: At first, we need an Android studio to work with Android programming. When we launched Android Studio after installation, we created a new project named TourMate.

Step 2: After creating the Android project, we designed the user interface. We

have completed the front-end UI using the XML language. There were many user interfaces, such as home page, log-in, signup, user profile, change information, and user input for recommending the suitability of the best place, the best local places, hotels, transportation services, travel agencies, and various activities to do. The user interface of our app consists of all the elements that travelers can view and use. We utilized Android's pre-built UI elements, including structured layout objects and UI controllers, to create the graphical user interface for our app.

Figure 3.5 shows the UI design of our Homepage.

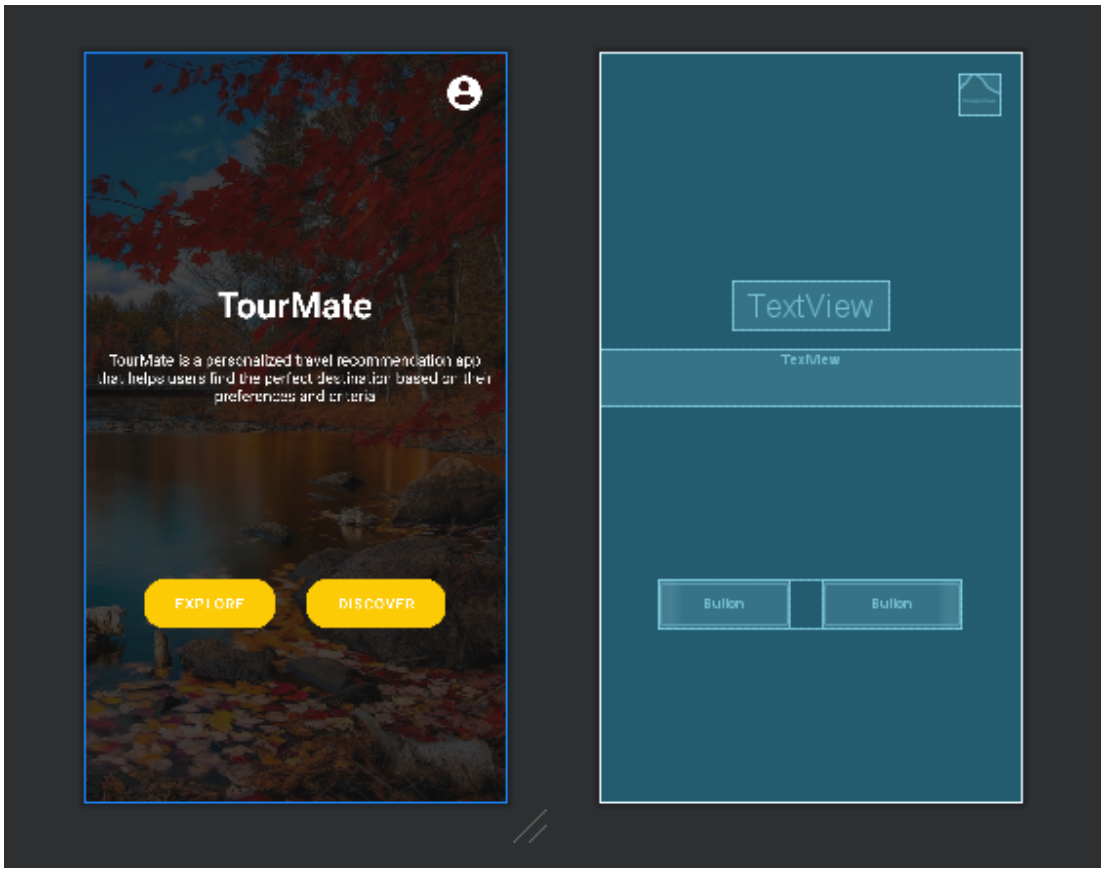


Figure 3.5: UI design of homepage

In UI design, in the upper right corner, there is the logged-in Traveler profile. At the center of the design are two buttons. One of them is Explore button and another one is the Discover button.

Figure 3.6 shows the UI design of our questionnaires page after clicking on the Discover button.

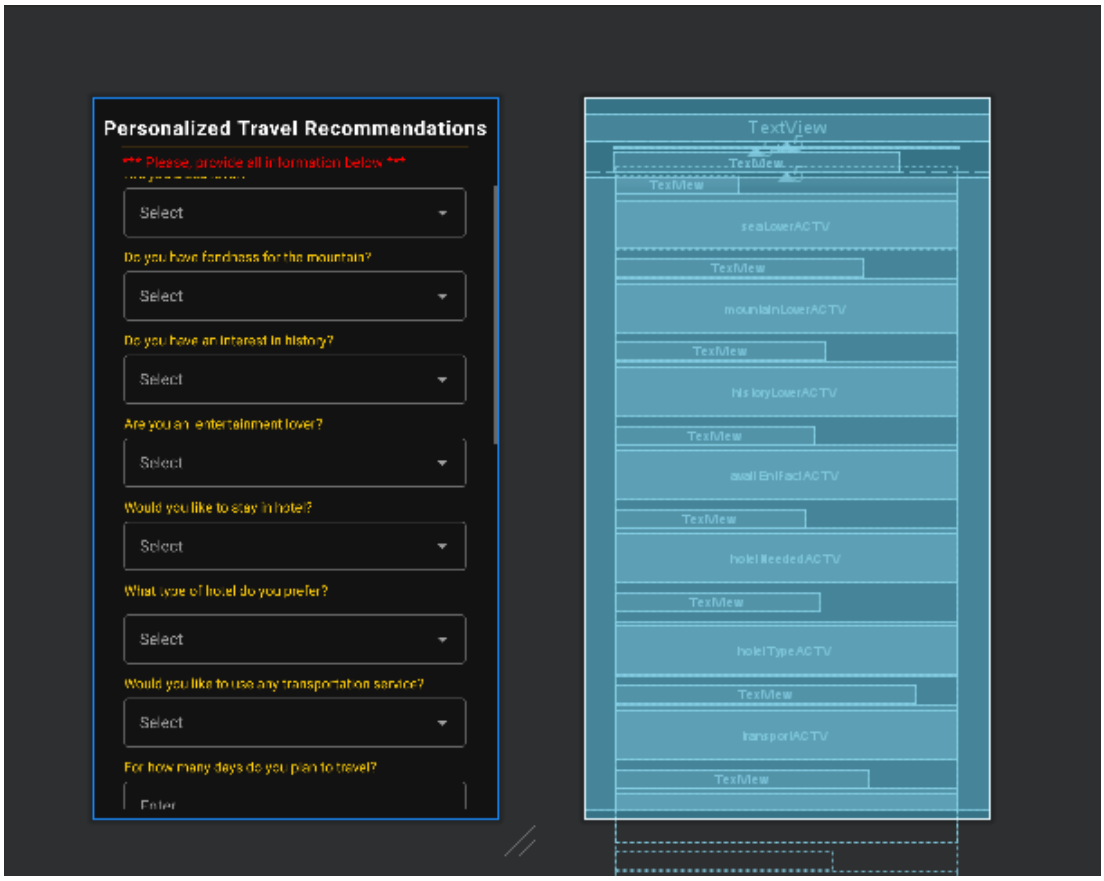


Figure 3.6: UI design of Discover button page

In this UI design, there are two text views in the upper which display the text "Personalized Travel Recommendations" and another text "**Please provide all information below**". There are a total of seventeen questions. In the lower portion, there is a predict button. The figure represents the UI design of a recommendation of a tourist place suitability on Android. This is the UI for showing the results of our system.

Figures 3.7 and 3.8 show the UI design of the result page.

In this UI design, there is one text view that displays the recommended suitability of the best place for travelers. And there is also a card view that shows the message "Best Wishes! This place is an excellent fit for your travel needs" when a place is suitable for a traveler. Otherwise, It shows the message "Sorry! This place may not be the best fit for your travel needs".

Step 3: In this step, we have created a database for our app. We utilized Firebase Real-Time Database to build the database. The stored database in Firebase is

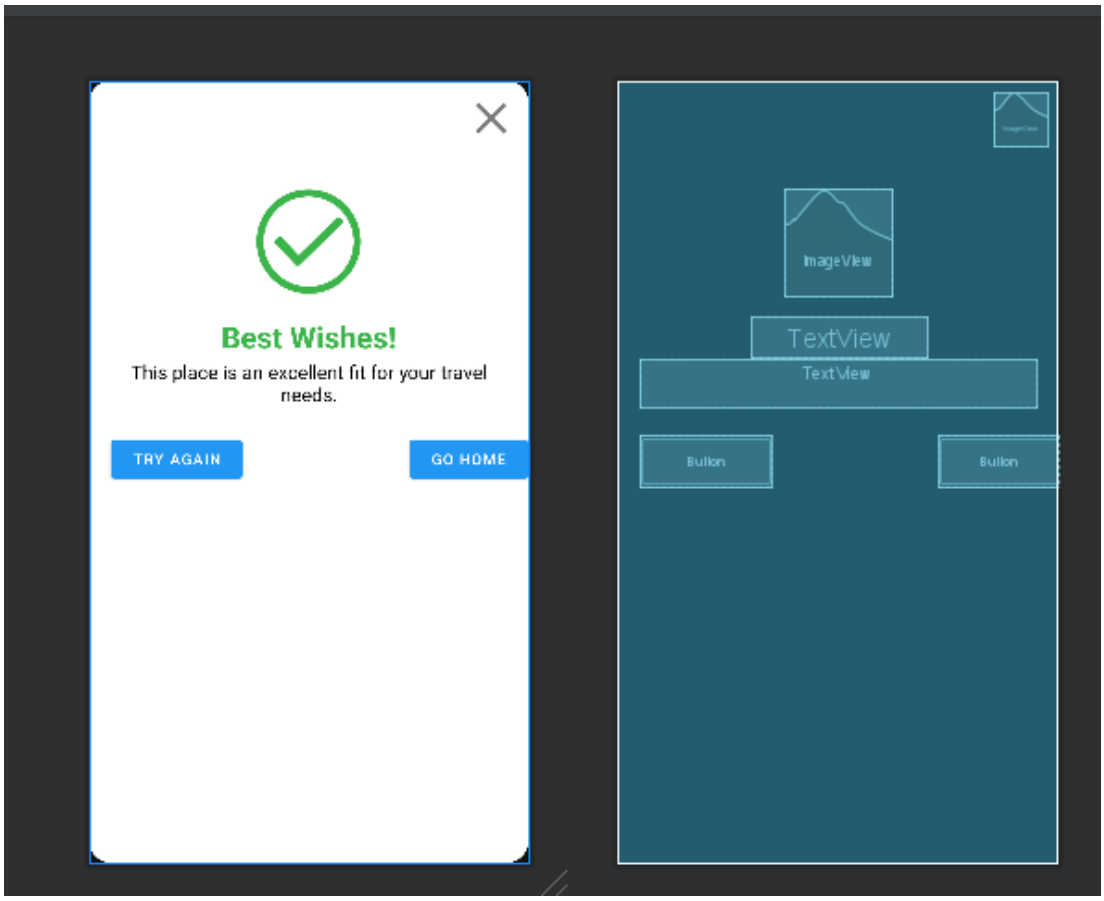


Figure 3.7: UI design of result page (a)

shown in Figures 3.9 and 3.10.

Google has also built a platform called Firebase for building mobile and web applications. A cloud-hosted NoSQL database called Firebase Real-Time Database allows users to sync and save data in real time. To provide users with notice, we used Firebase Cloud Messaging (FCM) multicast, which can transmit messages to numerous devices. Firebase Authentication is used to provide quick and easy developer authentication for the Realtime Database, which utilizes a declarative security paradigm to grant access based on user identity or data pattern matching. Additionally, Firebase serves as the storage for the required data. Figure 3.9 represents the user authentication database. We only need to obtain the user's authentication credentials from them and then provide them to the Firebase Authentication SDK in order to authenticate our users. This login information includes an email address, password, mobile number, user name, applicant occupation, etc. After receiving the credentials, Firebase will check them, and the user will then receive a response informing the user of the success or failure of the

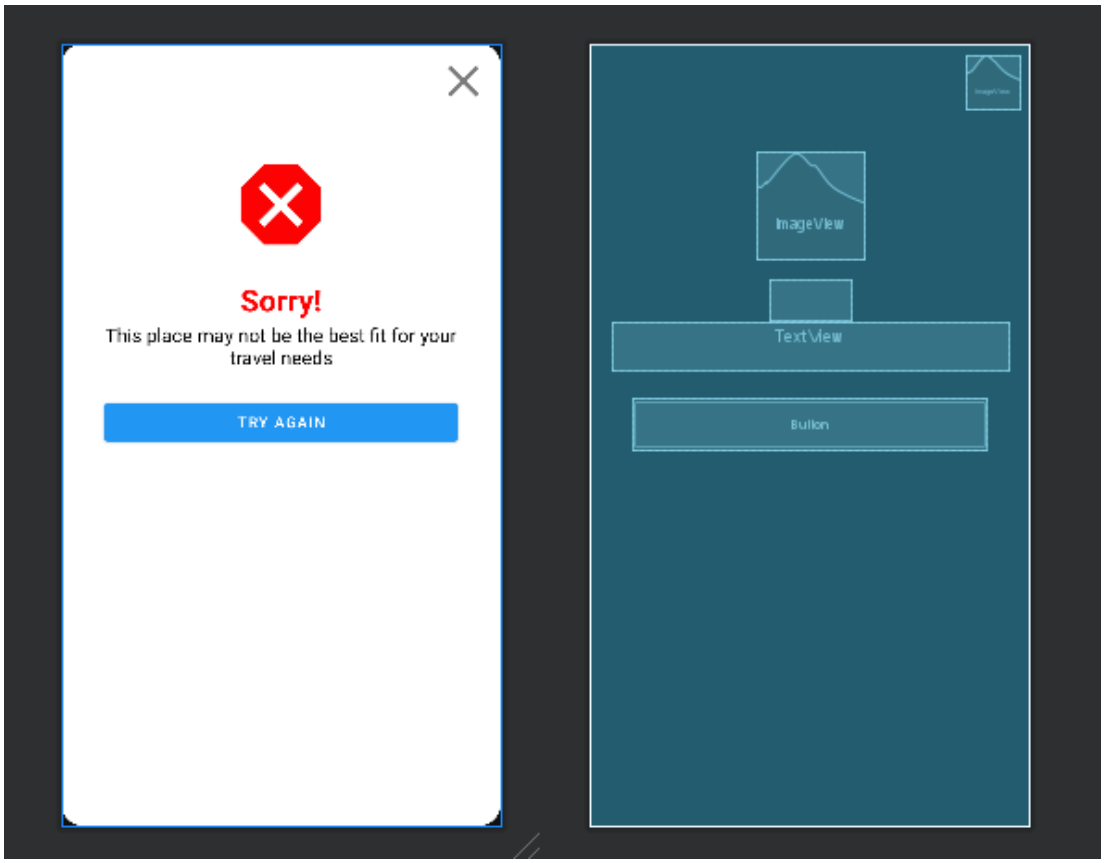


Figure 3.8: UI design of result page (b)

authentication. To create a user with an email address and a password, use the createUserWithEmailAndPassword() method. The arguments for this function are an email address and a password, which are validated before a new user is created. On the Firebase website, in the Authentication area, then on the Users page, our user will be registered, and we can view the registered user there. Users will encounter an error if they attempt to register several times using the same email address. We have a method called "signInWithEmailAndPassword()" that allows users to sign in with their email and password. This technique takes an email address and a password as arguments validates them, and then, if the validation is successful, signs the user into our application. Figure 3.10 represents the users' database. The tour-related description database and the recommended place suitability information database are in the firestorm of Firebase. In the tour-related description database, there are data on local places are stored.

Step 4: After designing our UI, we implemented the necessary functionality to handle user interactions and data input from the front end in a Java file.

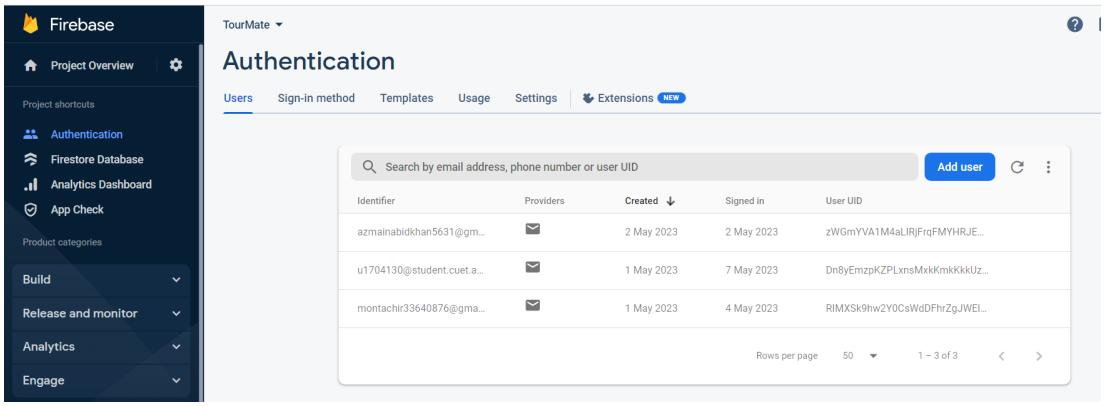


Figure 3.9: User authentication database

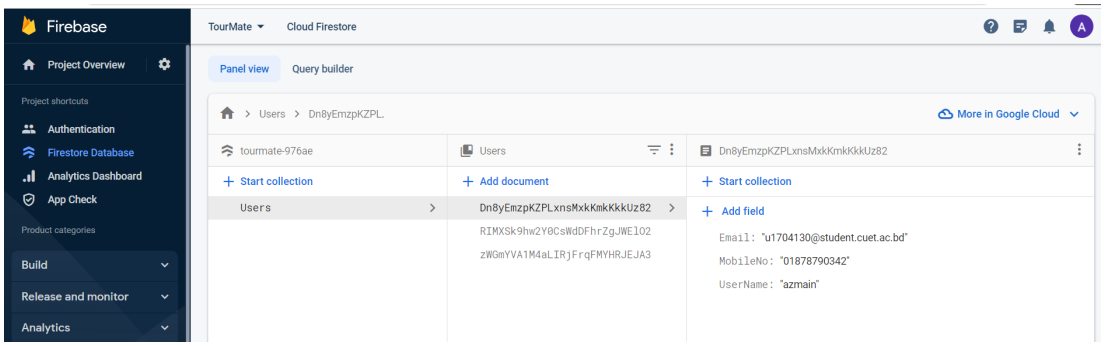


Figure 3.10: User database

Step 5: First, we saved our machine-learning model as an a.pkl file. Then we have the flask API using Python in Pycharm. Furthermore, when a traveler submits a post request, the Flask API will receive their data and send it to the machine learning model so that it can predict the output class. We will send the anticipated class as JSON to the Android app.

Step 6: the Figure represents the deployment of the ML model on Android. The API we developed was running locally on our system, there was a problem. Therefore, we had to launch our API online, and Railway was used to do so. One library called Volley was required to hit API.

Step 7: Finally, the user will fill in the required input for the suitability check. Then the input will be passed in JSON form through the API flask to the ML model, which has been saved as an a.pkl file. Then the suitability of the place will be checked by the saved ML model. The result will be shown in the mobile app and it will recommend the suitability of the place based on the user demand.

3.3 Conclusion

This chapter presents the proposed approach, which consists of two parts: the first part focuses on developing a tourist place suitability recommendation system for Bangladesh, while the second part focuses on providing personalized recommendations for local places based on user preferences. To predict the stage of best tourist place suitability, the best classification model is utilized. After predicting the best tourist place suitability, we also recommend the best local places for travelers according to their demands via mobile applications. The next chapter will show experimental findings as well as other information and some discussion.

Chapter 4

Results and Discussions

4.1 Introduction

A thorough justification of our method was given in the previous chapter. This chapter aims to demonstrate the comparative analysis of the model we have suggested. Also mentioned will be further research. This proposed method for determining the best tourist place suitability for travelers in Bangladesh was implemented in a Google Colab notebook using Python. For Android applications, we have used the Java language, which is used for developing the front-end in Android Studio, and the Firebase real-time database used for TourMate databases.

4.2 Dataset Description

In this project, we analyzed different reasons to determine if a certain place is suitable or not for a traveler which we explained in Chapter 3. For this, we have created a dataset by collecting answers from Google Forms. We asked travelers and travel agencies to provide us with tour-related information. In total, we have got the response of 868 data which is the size of our dataset. Figure 4.1 shows the features of the dataset.

Here we can see 5 are object-type data. Some of the sample data from the dataset is shown in Figure 4.2. An analysis of the importance of features in the dataset is also done, It is shown in Figure 4.3.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 868 entries, 0 to 867
Data columns (total 28 columns):
 #   Column                                                Non-Null Count  Dtype
---  ------                                                --------------  -----
 0   Timestamp                                             868 non-null    object
 1   Email                                                 868 non-null    object
 2   Is that place right for a sea lover?                  868 non-null    int64
 3   Is that place right for a man who loves mountains?    868 non-null    int64
 4   Is that place right for a history lover?              868 non-null    int64
 5   Available Entertainment Facilities                    868 non-null    int64
 6   Hotel needed                                          868 non-null    int64
 7   Transport needed                                      868 non-null    int64
 8   Days                                                  868 non-null    int64
 9   Suitable trip                                         868 non-null    int64
 10  Place                                                 868 non-null    int64
 11  Budget                                                868 non-null    int64
 12  Travel guide                                          868 non-null    int64
 13  Hotel type                                            868 non-null    int64
 14  Is the place full of natural and man-made attractions? 868 non-null   int64
 15  Traveling partner                                     868 non-null    int64
 16  Is the place safe for a tourist?                      868 non-null    int64
 17  Transport type                                        868 non-null    int64
 18  Hotel cost                                            868 non-null    int64
 19  Are there enough food stalls to enjoy various types of foods? 868 non-null int64
 20  Are the local people friendly with tourists?          868 non-null    int64
 21  Food cost                                             868 non-null    int64
 22  Transport_cost                                        868 non-null    int64
 23  famous Places                                         867 non-null    object
 24  Foods                                                 868 non-null    object
 25  Restaurants                                           868 non-null    object
 26  Your starting point                                   868 non-null    int64
 27  Ratings                                               868 non-null    int64
dtypes: int64(23), object(5)
memory usage: 190.0+ KB
```

Figure 4.1: Features of dataset

	D	E	F	G	H	I
1	Is that place right for a m	Is that place right for a hi	Available Entertainment f	Hotel needed	Transport needed	Days
16	1	0	1	0	1	1
17	1	0	1	1	1	1
18	1	0	1	1	1	2
19	1	0	1	1	1	2
20	1	0	1	1	1	2
21	1	0	1	1	1	2
22	1	0	1	1	1	2
23	1	0	1	1	1	1
24	1	0	1	1	1	2
25	1	0	1	1	1	1
26	1	0	1	1	1	2
27	1	0	1	1	1	2
28	1	0	1	1	1	2

Figure 4.2: Sample portion of dataset

4.3 Impact Analysis

In this research, we aimed to create a model that could determine whether or not a place was suitable for travelers. Finding a model that simply predicts suitability, however, is insufficient. As a result, we developed a mobile application for the Android operating system that can recommend local places, hotels, transportation services, travel agencies, and various activities to do for travelers in Bangladesh. We can infer that this project has implications for society and morality.

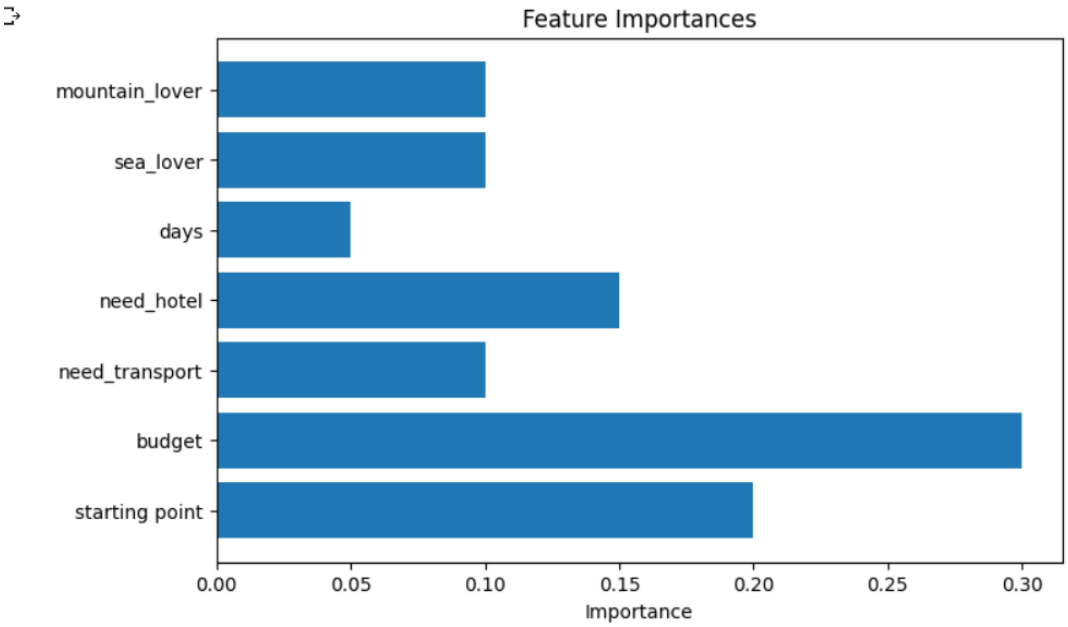


Figure 4.3: Comparison of features in order of importance

4.3.1 Social and Environmental Impact

This project will have a significant impact on society, particularly on travelers in Bangladesh. In Bangladesh, the tourism industry has a significant impact on the economy. It can earn a large amount of money from international tourists. There are many opportunities to boost the tourism industry in Bangladesh. Here people are interested in traveling and exploring new places. But they have no clear concept of the travel destination, time, duration, etc. Sometimes they have no idea that there are some places nearby. So they can't get the scope of traveling of these places. An Android-based smartphone application is being proposed here. It will help travelers to introduce a personalized travel recommendation system in Bangladesh. This app will save the valuable time of travelers by suggesting the suitability of their desired place. Besides it will recommend local places, hotels, transportation services, travel agencies, and various activities to do for travelers in Bangladesh. I have collected some user ratings for this TourMate app from my friends. These ratings are shown in Fig. 4.4.

From this user rating graph, it is seen that the good rating of every feature is high. And the number of medium ratings, no comments, and bad ratings descends in that order. The number of bad ratings is too low comparatively. The login and

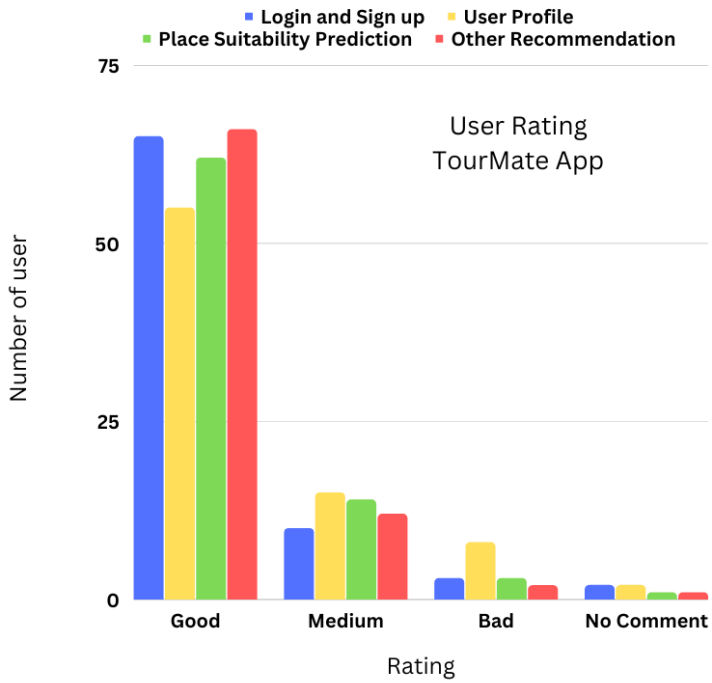


Figure 4.4: User Rating

signup features get the highest good ratings.

4.4 Evaluation of Framework

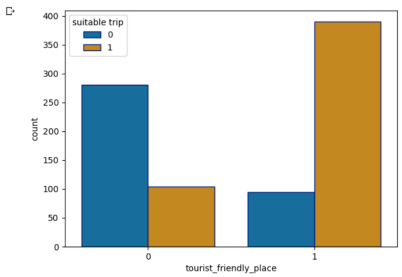
The outcomes of the experiment will be presented in this section. The validation results of the project are displayed in this crucial part.

4.4.1 Data Visualization

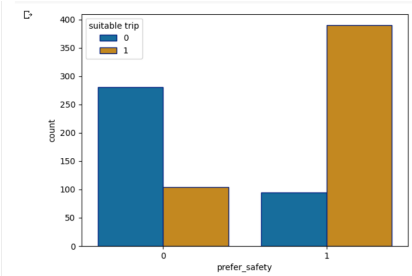
Because it helps people understand the relevance of data through visual representation, data visualization is crucial. Important co-relationships between data may also be seen through visual representation. As a result, displaying material in a visual manner is the ideal technique to comprehend it better.

Figure 4.5 represents the possibility of recommending the suitability of a place against tourist friendly place and safety. It is found from the count vs tourist friendly place graph and from the count vs prefer safety, which denotes the suitability status against tourist-friendly places, that if the traveler prefers a friendly

place and safety then the possibility of recommending suitability of a place is much higher than if they prefer a non-friendly place and non-safety.

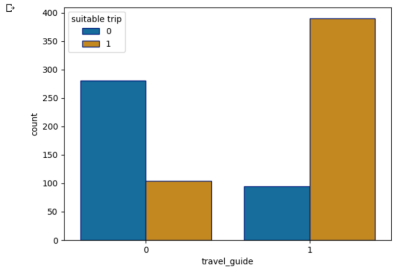


(a) Suitable trip against tourist friendly place

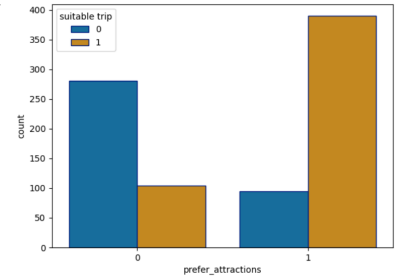


(b) Suitable trip against safety

Figure 4.5: Suitable trip data visualization based on tourist friendly and safety

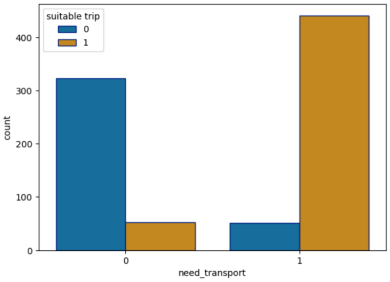


(a) Suitable trip against preferred travel guide

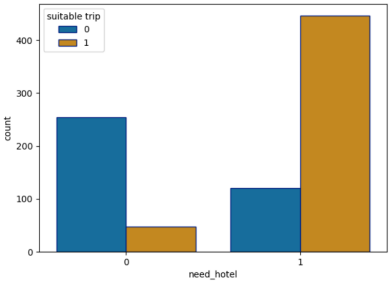


(b) Suitable trip against preferred attractions

Figure 4.6: Suitable trip data visualization based on travel guide and attractions



(a) Suitable trip against the need for transport



(b) Suitable trip against the need for hotel

Figure 4.7: Suitable trip data visualization based on transport and hotel

Figure 4.6 represents the possibility of recommending the suitability of a place against a travel guide and preferring attractions.

Figure 4.7 represents the possibility of recommending the suitability of a place against the need for transport need of a hotel.

4.4.2 Best Parameter for K-Nearest Neighbor Algorithm

For the best accuracy of the model, we have selected the optimal number of neighbors. In our model, we have found the best result for `n_neighbour=4`. The optimal number of neighbors was 4. Figure 4.8 shows that when the number of neighbors is 4, then the accuracy result is best.

```
plt.show()
```

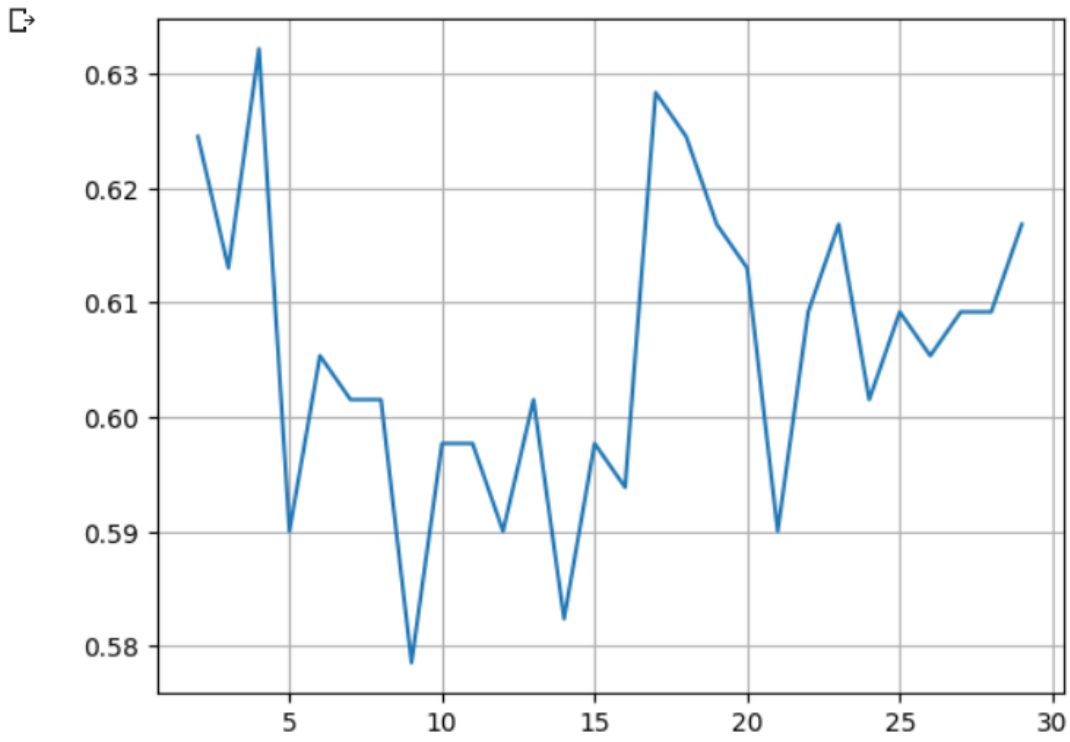


Figure 4.8: Model accuracy comparison against neighbors

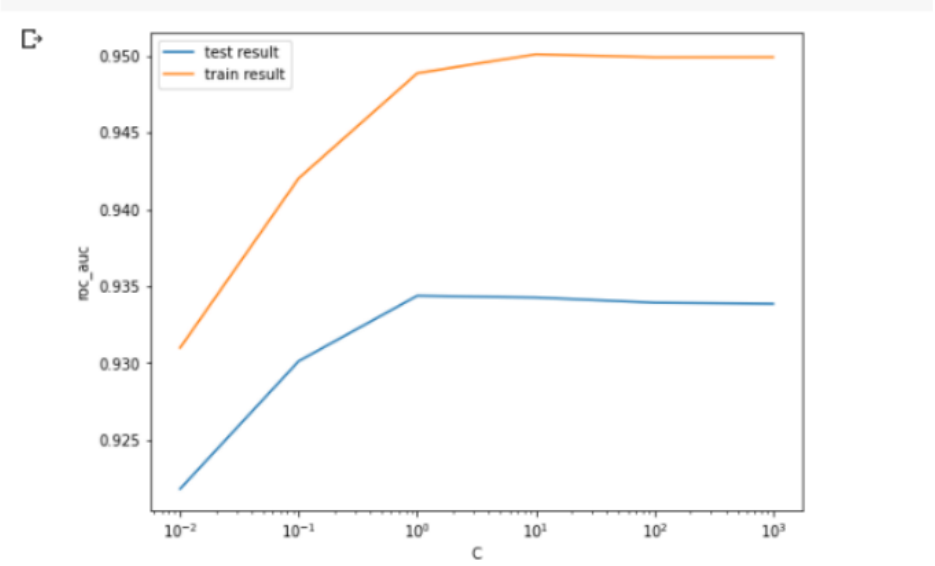
4.4.3 Best Parameter for Random Forest Algorithm

Figure 4.9 represents the best parameter for the Random Forest Algorithm in our model. The best parameters, such as:

- `bootstrap`: True
- `max_features`: 'auto'
- `min_samples_leaf`: 1
- `min_samples_split`: 5
- `n_estimators`: 30

4.4.4 Best Parameter for Logistic Regression Algorithm

For the best accuracy of the model, we have selected the optimal number of C. In our model, we have found the highest test `roc_auc` is 0.936754367812 at $C = 1$. The optimal number of C was 1. Figure 4.9 shows the `AUC_ROC` curve for $C = 1$.



(a) `AUC_ROC` curve for Logistic Regression

```
[49] rf_grid_model.best_params_

{'bootstrap': True,
 'max_features': 'auto',
 'min_samples_leaf': 1,
 'min_samples_split': 5,
 'n_estimators': 30}

[50] rf_grid_model.best_estimator_

RandomForestClassifier(max_features='auto', min_samples_split=5,
                       n_estimators=30)
```

(b) Best parameter for Random Forest Algorithm

Figure 4.9: Logistic Regression and Random Forest Algorithm

4.4.5 Mobile Application Framework

Figure 4.12 represents the home page of the TourMate mobile app.

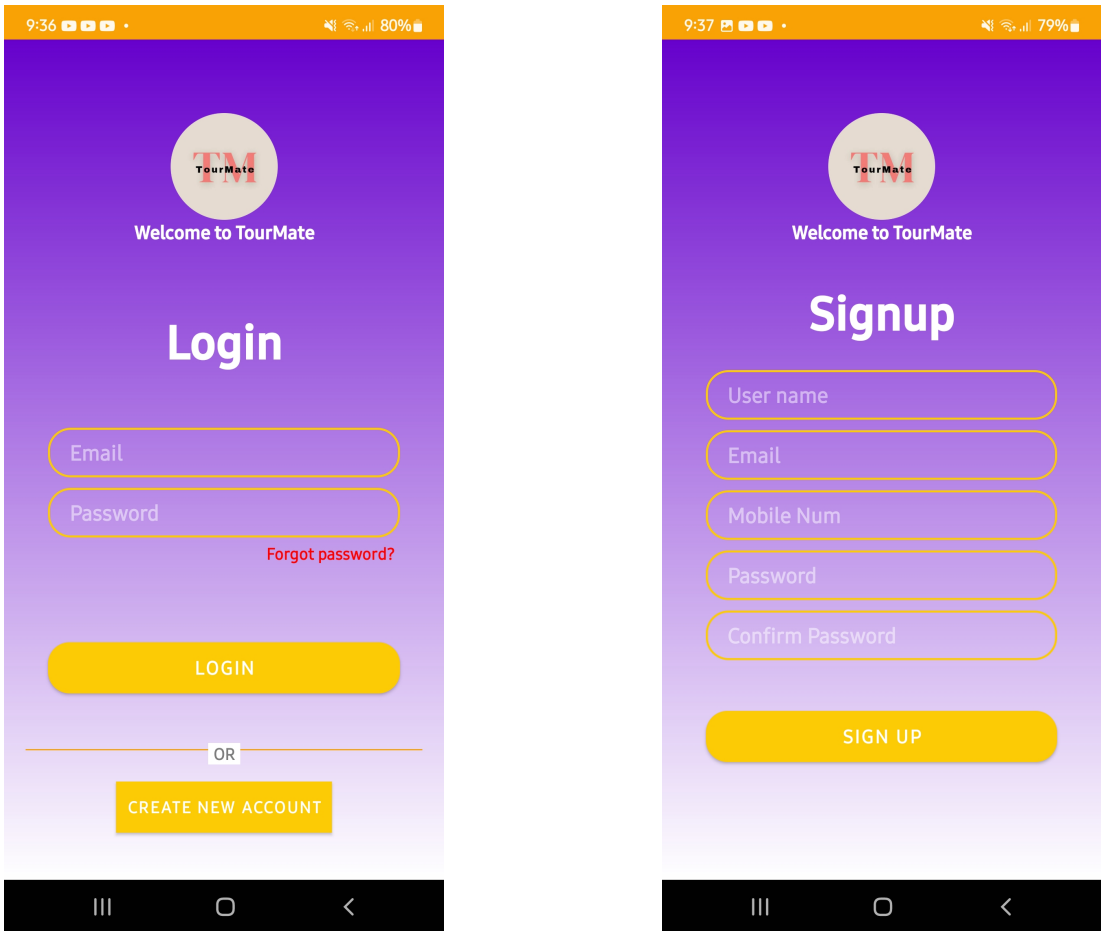


Figure 4.10: Login and Sign up page

The user profile is in the upper right corner of the home page. In the middle section, there is a Discover module where travelers can check the suitability of their desired place. Traveler can access their profile by clicking the user profile section. There is also an Explore section where the users will answer some questions to see the best local places.

Figure 4.10 represents the user signup and login interface. The traveler can enter the home section immediately if he/she has previously signed in to the app. If not, he/she must type the login phrase. This app requires an email address and password for access. The home task interface is shown if the email is verified. If the traveler is new to this TourMate app, he/she must first register. A username, email address, mobile number, password for your occupation, and a confirmed password are needed to register for this app. The home activity interface is shown if the password matches the confirmed password.

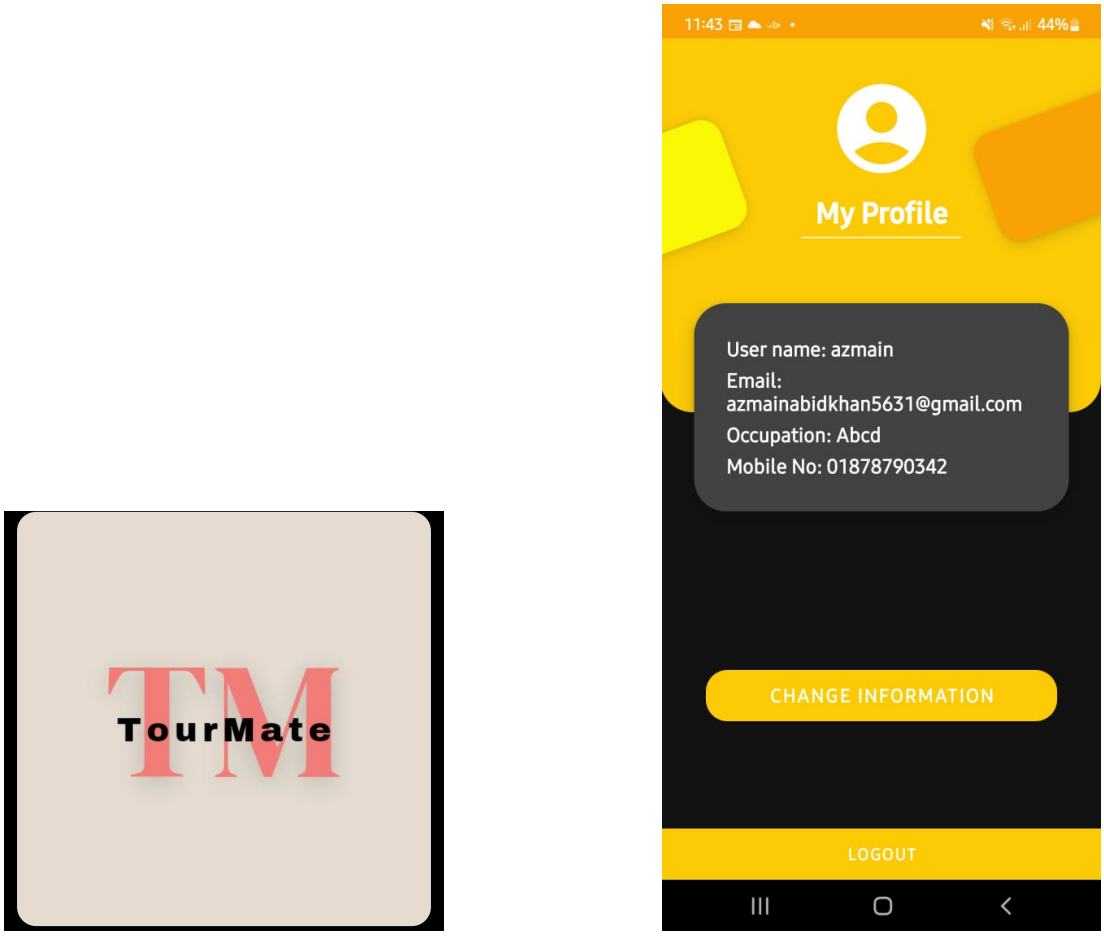


Figure 4.11: Logo and profile section

Figure 4.11 shows the logo of the app and the user profile. After choosing a user profile, a traveler can view and edit the information on their profile. Travelers are able to see their personnel information in the User Profile interface, and they can change their information if required. They can also log out of the system if they want.

The interface for recommending the suitability of the desired place is shown in Fig. 4.13. With the necessary inputs, travelers can determine their place suitability. They must enter the necessary information, including the required inputs such as whether they are Sea lovers, mountain lovers, history lovers, entertainment lovers, or not. Besides, Do they need any hotel, what type of hotel need, need any transport service or not must be included. Finally, the duration of the trip, desired place, budget, require any travel guide or not, do they prefer attractions, having any traveling partner or not, do they prefer safety, whether are they foodies

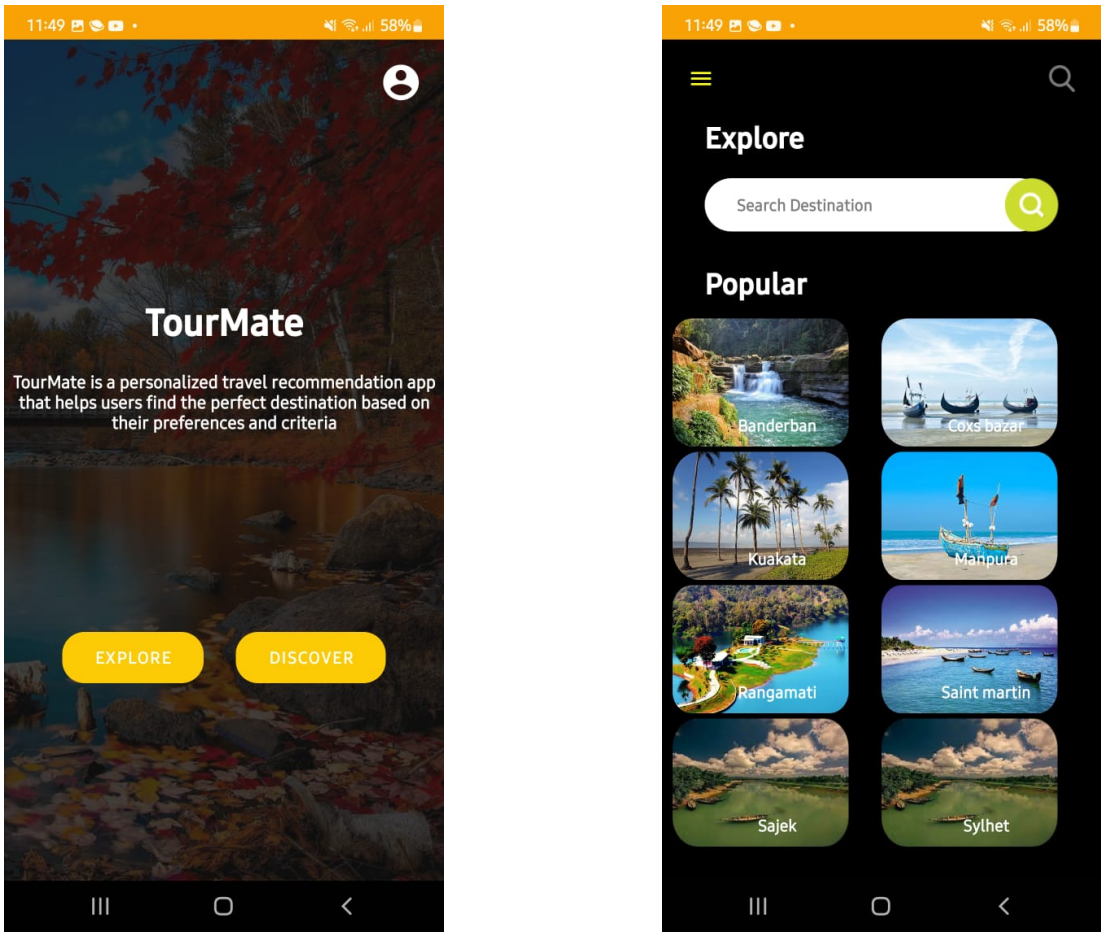


Figure 4.12: Homepage section

or not, do they prefer a tourist-friendly place or not, their starting point must also be given. Here, a machine learning model is used to determine suitability.

Figure 4.14 shows the output of the model. If the user's desired place is suitable then it will show a message: "Best Wishes! This place is an excellent fit for your travel needs". Then the user can click on the "Try Again" option for further check or can click the "Go Home" option for checking other factors. Otherwise, it will show that: "Sorry! This place may not be the best fit for your travel needs". After that, the user can click on the "Try Again" option.

4.5 Evaluation of Performance

The performance of a prediction model is evaluated using a confusion matrix, which shows the number of accurate and inaccurate predictions for each class. The

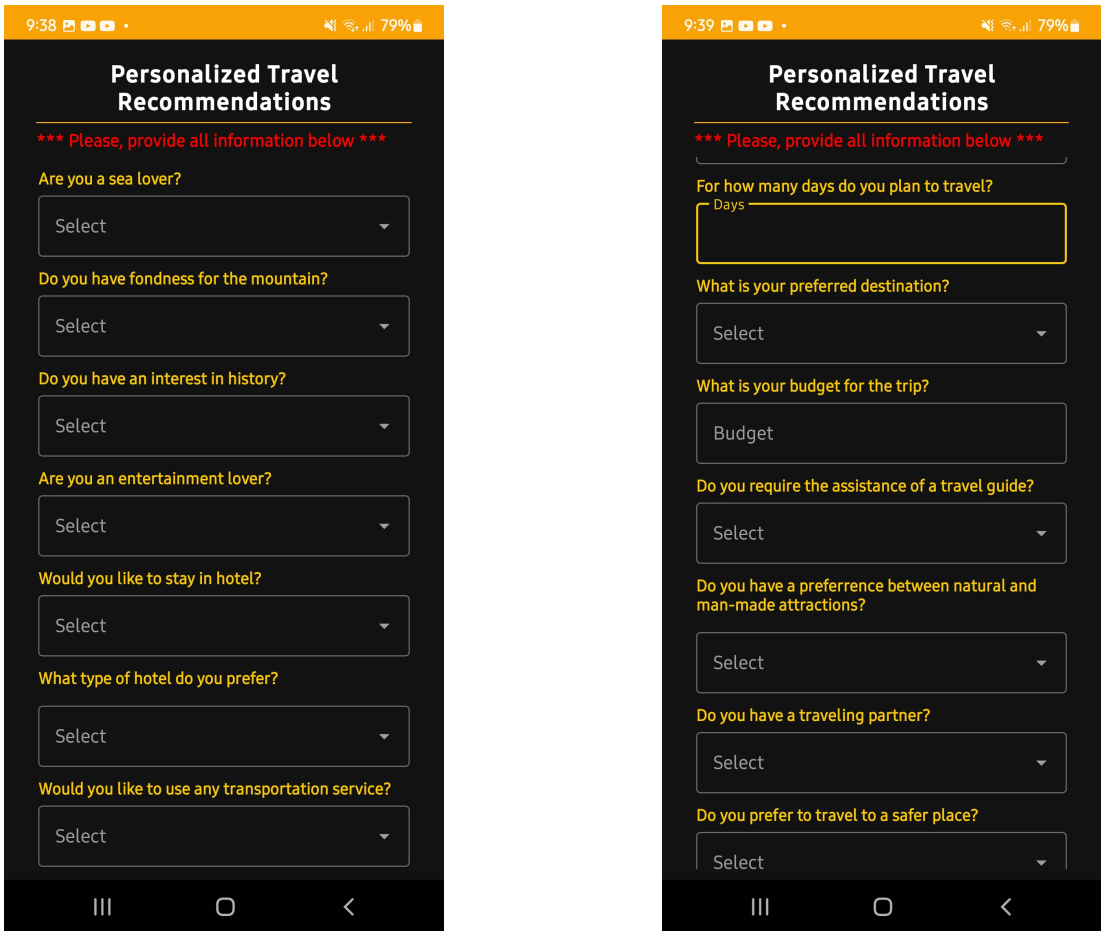


Figure 4.13: Gathering information from travelers for checking place suitability

key evaluation metrics for a classification problem are accuracy, precision, recall, specificity, and F1-score, which are calculated using the following equations:

	Predicted Not Eligible	Predicted Eligible
Actual Not Eligible	True Negative	False Positive
Actual Eligible	False Negative	True Positive

Table 4.1: Confusion Matrix

$$\text{Accuracy} = \frac{TruePos+TrueNeg}{TruePos+FalsePos+FalseNeg+TrueNeg}$$

$$\text{Recall} = \frac{TruePos}{TruePos+FalseNeg}$$

$$\text{Precision} = \frac{TruePos}{TruePos+FalsePos}$$

$$\texttt{F1\_score} = 2 \times \frac{Precision \times Recall}{Precision+Recall}$$

TruePos: TruePos (True Positive) indicates the number of tourists who have been correctly identified as a suitable place to visit.

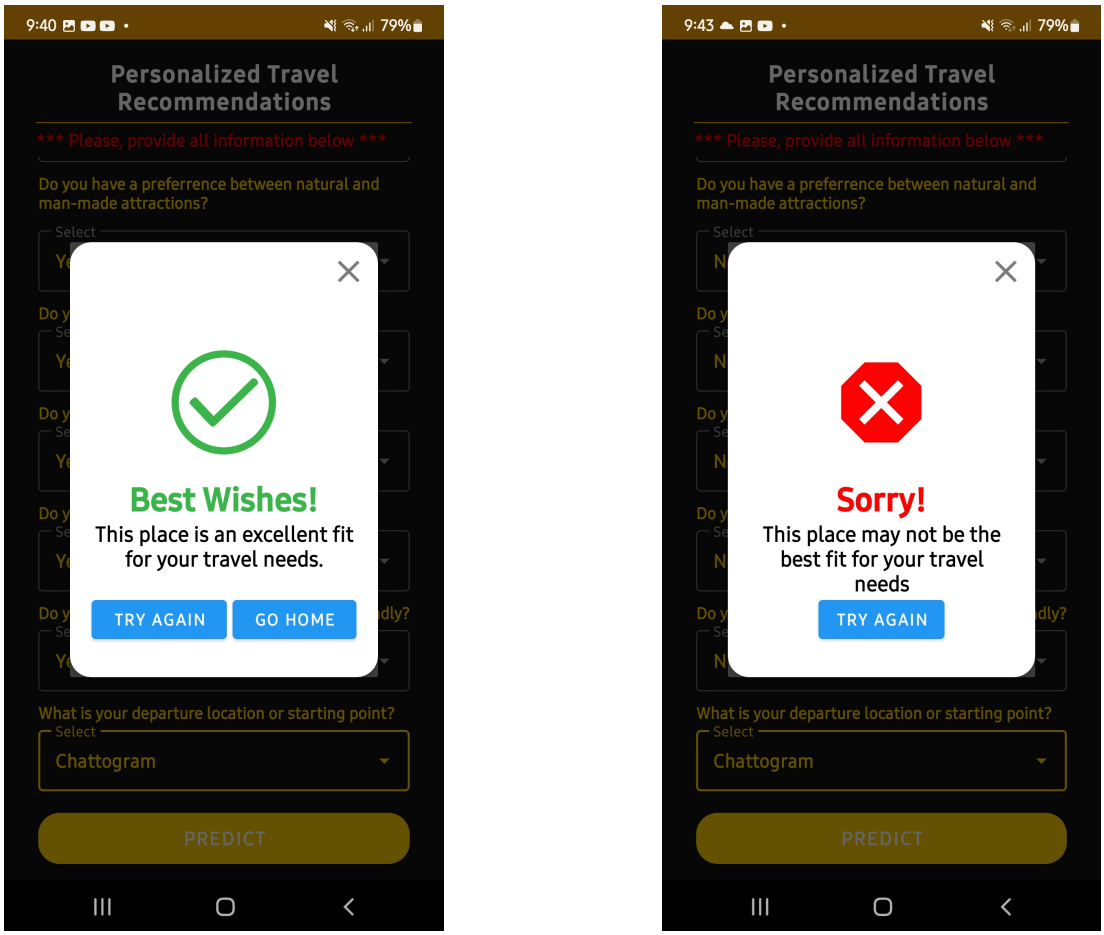


Figure 4.14: Place suitability is displayed in the mobile app

TrueNeg: TrueNeg (True Negative) is a measure of the number of places that have been appropriately categorized but are not suitable for tourists to visit.

FalsePos: FalsePos is a measure of the number of places that were mistakenly labeled as suitable for tourists to visit when, in fact, they were not. A Type I mistake is often referred to as FalsePos.

FalseNeg: FalseNeg indicates the number of places that were mistakenly excluded as suitable for tourists to visit. A Type II mistake is often referred to as FalseNeg.

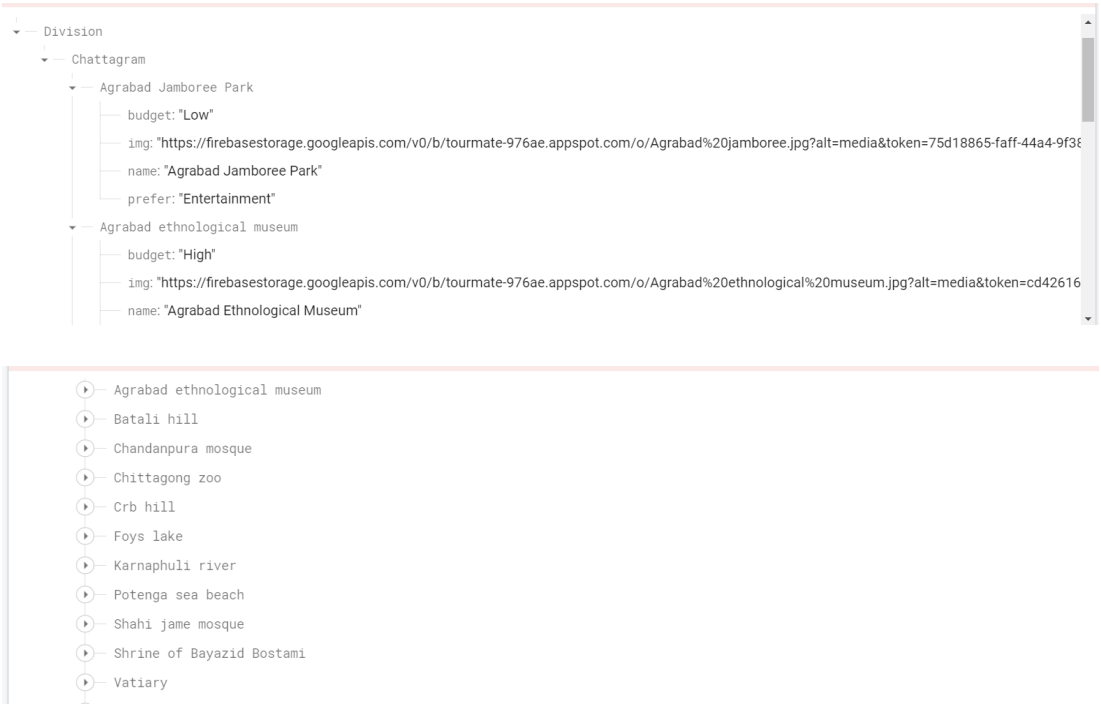


Figure 4.15: Firebase Real-Time Database

Serial No.	Method	Accuracy
01	K-Nearest Neighbor(KNN) Classifier Algorithm	63%
02	Logistic Regression Algorithm	91%
03	Random Forest Algorithm	98%
04	Decision Tree Classifier Algorithm	96%
05	Linear Regression Algorithm	74%
06	Naive Bayes Algorithm	91%

Table 4.2: The accuracy of several algorithms

4.5.1 Performance Evaluation Using Model Accuracy and Confusion Matrix

Table 4.2 represents the accuracy comparison of all six algorithms. The model accuracy for six different classification algorithms (K-Nearest Neighbor Classifier, Random Forest, Logistic Regression, Decision Tree Classifier, Linear Regression, and Naive Bayes) are 63%, 91%, 98%, 96%, 74%, and 91%, respectively, and the Random Forest algorithm has the highest accuracy. So, finally, the Random Forest classifier is selected as the predictive model of the proposed system giving the best accuracy among the six of these algorithms. So, finally, the Random Forest classifier is selected as the predictive model of the proposed system.

Table 4.3 also represents the performance comparison of all five algorithms.

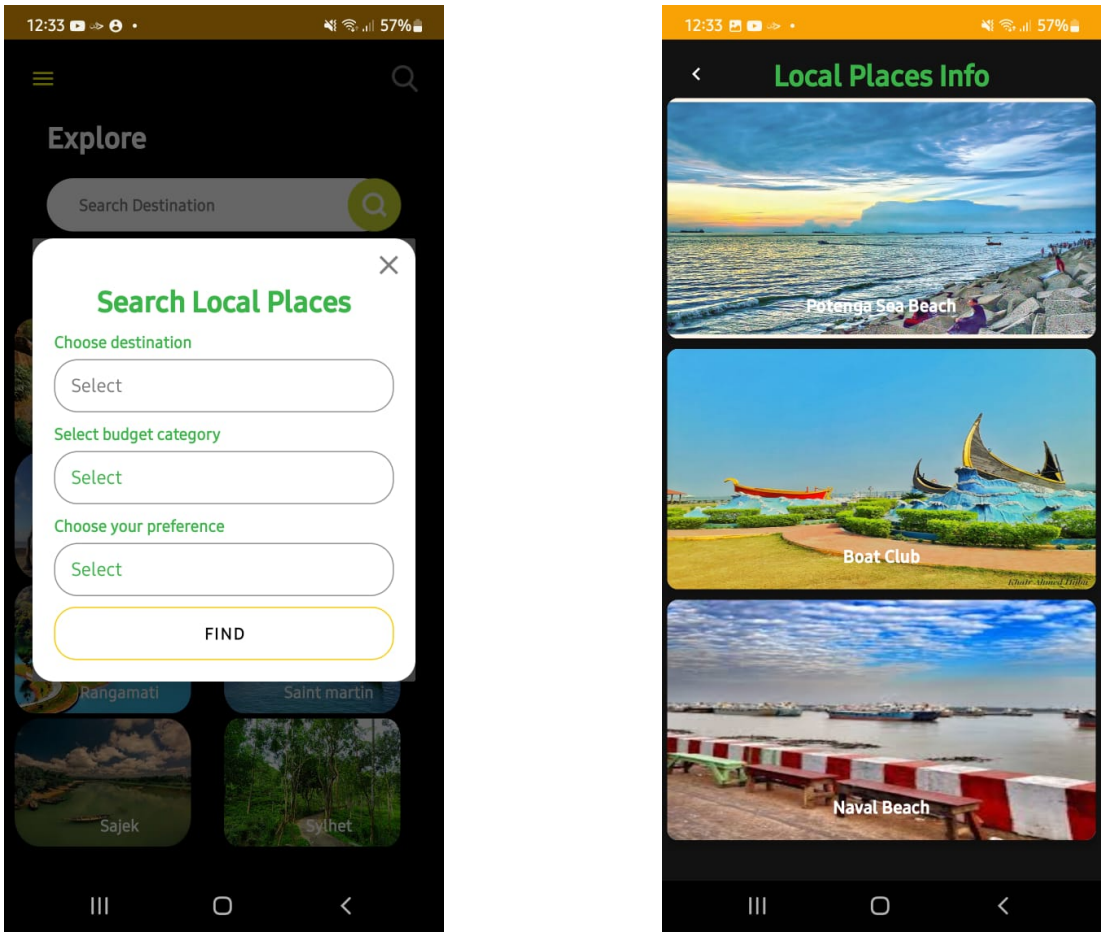


Figure 4.16: Recommendations of local places in the mobile app (a)

Precision: The precision metric measures the proportion of correctly predicted positive observations among all predicted positive observations. The precision values for K-Nearest Neighbor, Logistic Regression, Random Forest, Decision Tree, and Naive Bayes algorithms are 68%, 91%, 98%, 95%, and 95%, respectively. Among these algorithms, the Random Forest model achieves the highest precision.

Recall: Recall measures the proportion of actual positive observations that were correctly predicted, out of all positive observations. The recall values for K-Nearest Neighbor, Logistic Regression, Random Forest, Decision Tree, and Naive

Algorithm	Precision	Recall/Sensitivity	Specificity	F1-Score
K-Nearest Neighbor(KNN) Algorithm	68%	66%	59%	67%
Logistic Regression Algorithm	91%	93%	88%	92%
Random Forest Algorithm	98%	99%	97%	98%
Decision Tree Classifier Algorithm	95%	97%	97%	96%
Naive Bayes Algorithm	95%	89%	93%	92%

Table 4.3: The performance of several algorithms

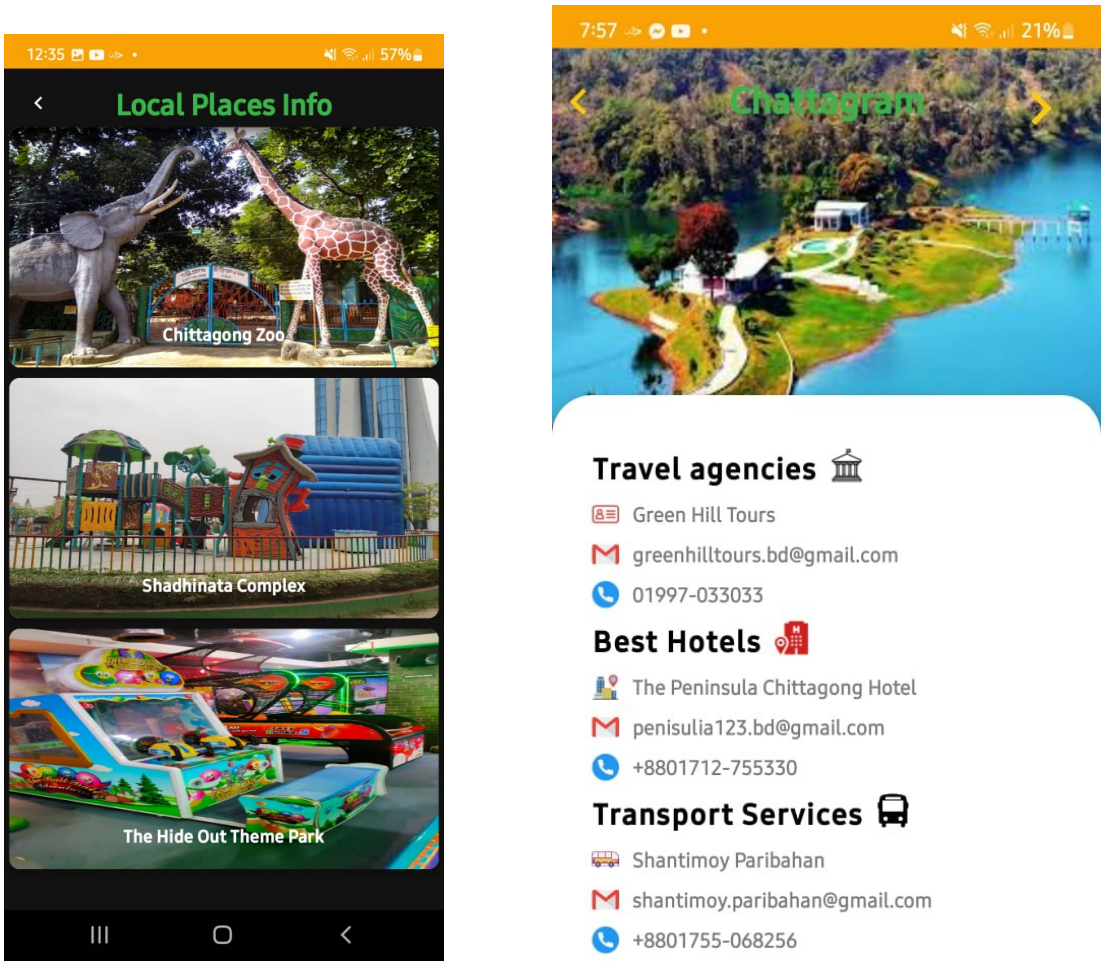


Figure 4.17: Recommendations of local places and others in the mobile app

Bayes algorithms are respectively 66%, 93%, 99%, 97%, and 89%. It can be observed that the Random Forest algorithm has the highest recall among these five algorithms.

Specificity: The percentage of true negatives that were projected as negatives is known as specificity (or true negative). This suggests that a further percentage of true negatives—which were formerly thought to be positive and could be referred to as false positives—will occur. The specificity for K-Nearest Neighbor, Logistic Regression, Random Forest, Decision Tree, and Naive Bayes algorithms are respectively 59%, 88%, 97%, 97%, and 93%. It is seen that the Random Forest algorithm gives the best specificity among the four of these algorithms.

F1-Score: The F1-score is a measure of a model's accuracy on a dataset and is commonly used to evaluate binary classification systems that classify examples

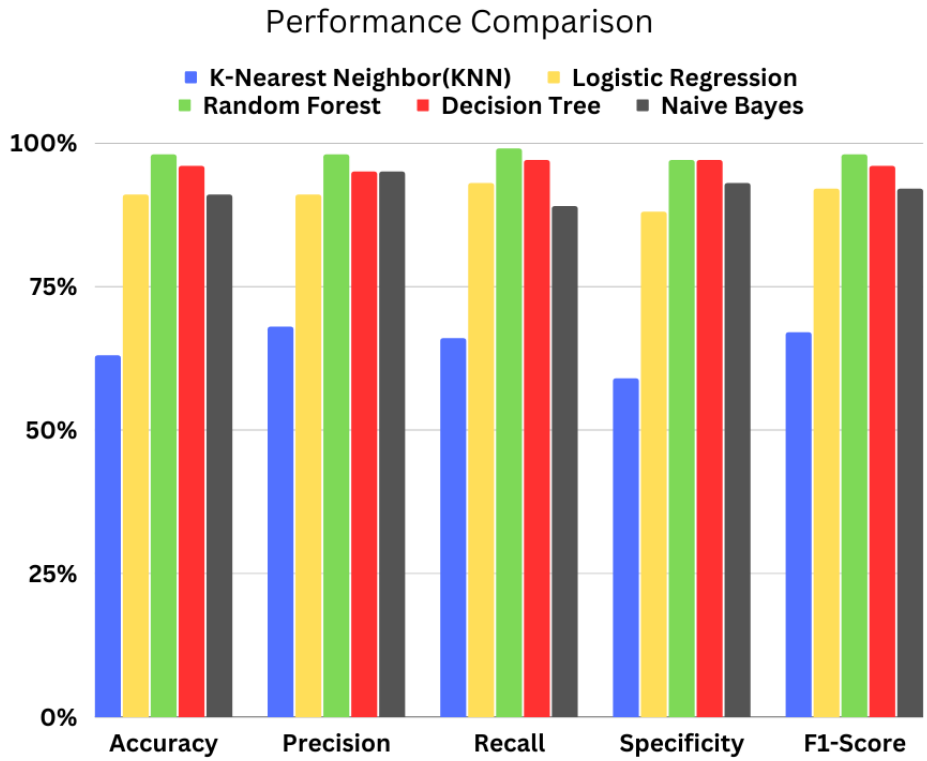


Figure 4.18: Performance Comparison

as "positive" or "negative." It is the harmonic mean of the model's precision and recall, providing a way to balance the two metrics. The F1-score for K-Nearest Neighbor, Logistic Regression, Random Forest, Decision Tree, and Naive Bayes algorithms are respectively 67%, 92%, 98%, 96%, and 92%. It is observed that the Random Forest algorithm outperforms the other algorithms in terms of the F1 score.

We may conclude that the model using the Random Forest method provides the best prediction result after reviewing the overall performance. For the purpose of predicting the suitability of the desired place for travelers, we have selected the model with the Random Forest algorithm.

Figure 4.19 represents the confusion matrix of the K-Nearest Neighbor Algorithm. The prediction model using this algorithm can predict the 100 places which have been correctly identified as suitable for travelers and the 66 places which have been appropriately categorized but are unsuitable for travelers. This the model predicted 45 places that were mistakenly labeled as suitable when, they were

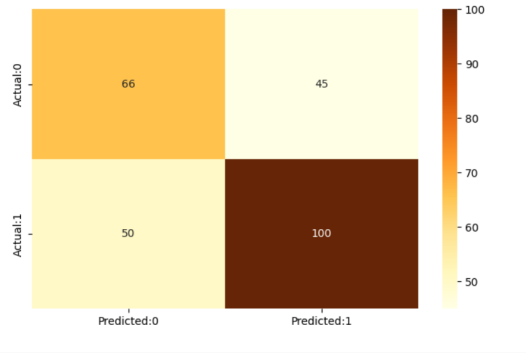


Figure 4.19: Confusion matrix of K-Nearest Neighbor Algorithm

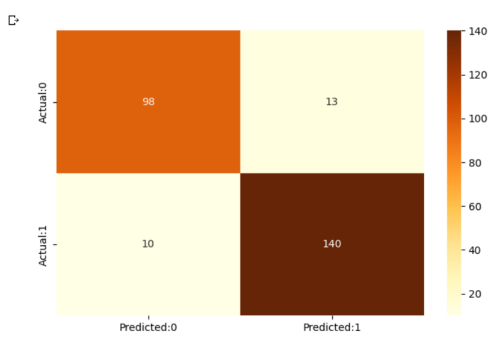


Figure 4.20: Confusion matrix of Logistic Regression Algorithm

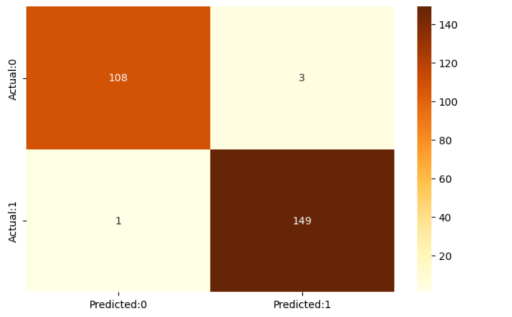


Figure 4.21: Confusion matrix of Random Forest Algorithm

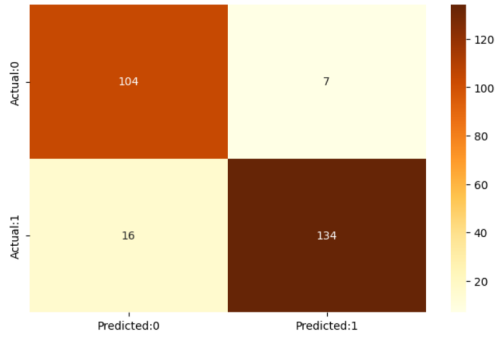


Figure 4.22: Confusion matrix of Naive Bayes Algorithm

not and also predicted 50 places which were mistakenly excluded from place suitability.

Figure 4.20 also represents the confusion matrix of the Logistic Regression Algorithm. The prediction model using this algorithm can predict the 140 places which have been correctly identified as suitable for travelers and the 98 places which have been appropriately categorized but are unsuitable for travelers. This model predicted 13 places that were mistakenly labeled as suitable when they were not and also predicted 10 places that were mistakenly excluded from place suitability. Figure 4.21 represents the confusion matrix of the Random Forest Algorithm. The prediction model using this algorithm can predict the 149 places which have been correctly identified as suitable for travelers and the 108 places which have been appropriately categorized but are unsuitable for travelers. This the model predicted 3 places that were mistakenly labeled as suitable when, they were not and also predicted 1 place which was mistakenly excluded from place

suitability.

Figure 4.22 also represents the confusion matrix of the Naive Bayes Algorithm. The prediction model using this algorithm can predict the 134 places which have been correctly identified as suitable for travelers and the 104 places which have been appropriately categorized but are unsuitable for travelers. This the model predicted 7 places that were mistakenly labeled as suitable when, they were not and also predicted 15 places which were mistakenly excluded from place suitability.

4.5.2 Performance Evaluation Using AUC-ROC Curve

A common tool for evaluating binary classification models is the Receiver Operating Characteristic (ROC) curve. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold values to distinguish between "signal" and "noise." The classifier's ability to accurately distinguish between classes is measured by the Area Under the Curve (AUC), which summarizes the ROC curve. A higher X-axis value on the ROC curve indicates a higher proportion of false positives compared to true negatives, while a higher Y-axis value indicates a higher proportion of true positives compared to false negatives. Selecting a threshold that balances the trade-off between false positives and false negatives is crucial for model performance. The equation of TPR and FPR are given below:

$$TPR = \frac{TruePositive}{TruePositive+FalseNegative}$$

$$FPR = \frac{FalsePositive}{TrueNegative+FalsePositive}$$

Where TPR denotes True Positive Rate. FPR denotes False Positive Rate. In this fig. 4.23, we can see that Random Forest provided the best performance as it occupies the curve at the far left and top.

The Area Under the Curve (AUC) values for the ROC curves of the K-Nearest Neighbor, Random Forest, Logistic Regression, Decision Tree, and Naive Bayes algorithms are 66%, 100%, 98%, 99%, and 97%, respectively. The AUC provides a measure of the model's ability to distinguish between positive and negative

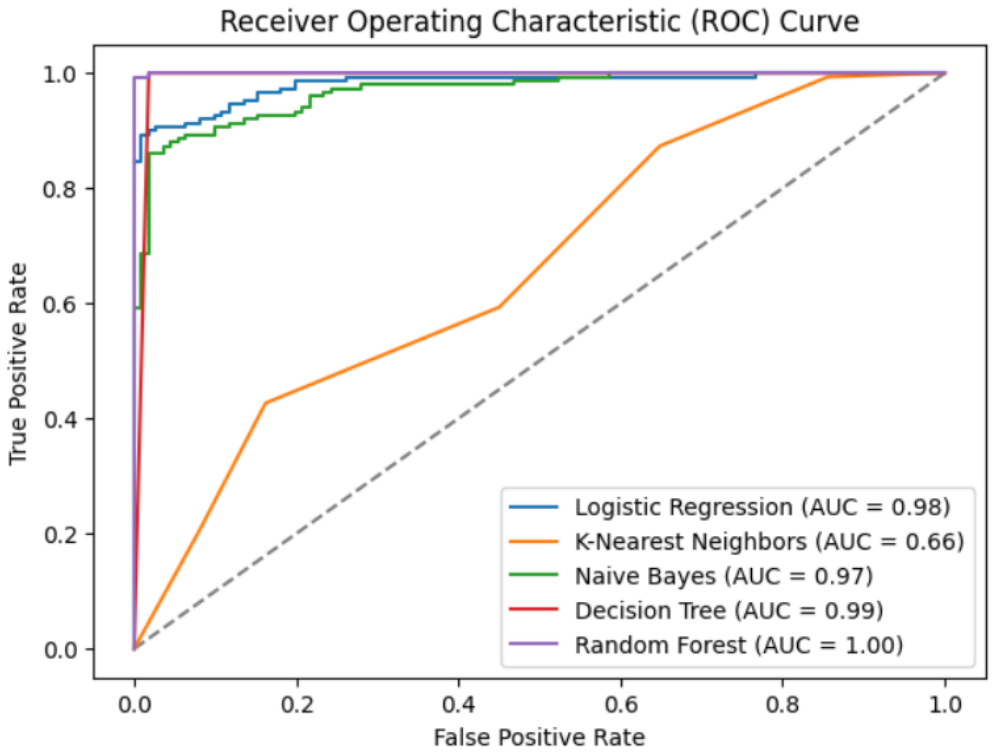


Figure 4.23: ROC-AUC curve of five algorithms model

classes, with a higher AUC indicating better performance. The ROC curve is a widely used tool in the machine learning community to compare the performance of different models. Among the five algorithms, the Random Forest algorithm achieved the best AUC score. A perfect model has an AUC score of 1.

4.5.3 Performance Evaluation for App Recommendation

In our model, the random forest algorithm gives us the best result compared with the other algorithms. The model can predict the suitability of travelers' desired places, and the result is displayed in the mobile app. For recommendation, they must enter the necessary information, including the required inputs such as whether they are Sea lovers, mountain lovers, history lovers, entertainment lovers, or not. Besides, Do they need any hotel, what type of hotel need, need any transport service or not must be included. Finally, the duration of the trip, desired place, budget, require any travel guide or not, do they prefer attractions, having any traveling partner or not, do they prefer safety, whether are they foodies or not, do they prefer a tourist-friendly place or not, their starting point must

also be given. If a particular place is suitable for a traveler, then the message "Congratulations! This place is an excellent fit for your travel needs" will be displayed. Otherwise, the message "Sorry! This place may not be the best fit for your travel needs" will be displayed.

4.6 Conclusion

The overall effectiveness of the suggested strategy has been reviewed in this chapter. The dataset's detailed explanation has been provided. Also demonstrated is the feature extraction procedure. Comparative results of several algorithms were also discussed. The module created for determining place suitability generates results based on many matrices, including precision, recall, specificity, f1-score, etc. A visualization component provides a clear view of the parameter changes. There are two parts to the entire system. The initial step is to create a travel suitability prediction model, followed by recommending the top local places to visit based on user data and the prediction model. We will provide a brief overview of the outcomes of these two steps.

Chapter 5

Conclusion

5.1 Conclusion

This thesis work aimed to develop a prediction model for determining the suitability of travel destinations for individual users based on their preferences and past travel experiences. For achieving my goal I have collected data on various factors that influence travel decisions, such as whether they are Sea lovers, mountain lovers, history lovers, entertainment lovers, or not. Besides, Do they need any hotel, what type of hotel need, need any transport service or not must be included. Finally, the duration of the trip, desired place, budget, require any travel guide or not, do they prefer attractions, having any traveling partner or not, do they prefer safety, whether are they foodies or not, do they prefer a tourist-friendly place or not, and their starting point. We then trained several machine learning algorithms, such as Decision Tree, Naive Bayes, K-Nearest Neighbor, Logistic Regression, and Random Forest, to predict the suitability of a place for a given user.

In terms of accuracy, recall, specificity, and F1 score we have seen that the Random Forest algorithm worked better. Besides, it had the highest AUC, indicating that it could better differentiate between true positives and true negatives than the other algorithms. That's why this algorithm predicts the suitability of the destination. Finally, this thesis work gives important insights to develop personalized travel recommendations for travelers. Here we have discovered and understood the complex factors that are necessary for travel decisions to predict suitability, by using machine learning algorithms. We hope that this is just a starting point for future research into the field of personalized travel recommendations. It will help to enhance the travel experience for travelers around the world.

5.2 Future Work

Some potential future work regarding my work could include:

Incorporating additional data sources: Here we utilized various data sources such as user reviews, travel agencies, and Facebook travel pages. But many other resources of data could be used to improve the accuracy of the prediction model and recommendations. For example, other social media data and user-generated data on travel websites could provide valuable insights into local trends and preferences.

Improving the user interface and user experience: There is also much significant potential to improve the user experience and also the user interface of the travel application. Example: Incorporating personalized travel planning tools, adding social sharing features, or generating more better navigation and search facilities.

Expanding the scope of the recommendation engine: Current recommendation engine focuses primarily on local places, hotels, transportation, travel agencies, and activities to do. But there is much potential to extend the scope of recommendations including other factors like cultural experiences and dining options to provide a more comprehensive travel planning experience for travelers.

Testing the model on different populations: The current prediction model and recommendation engine were developed using data from a specific population and geographic region. There is potential to test the model on different populations and geographic regions to determine whether it is applicable in other contexts and to identify any necessary adjustments.

Collaborating with industry partners: Here we have not collaborated with travel industry partners such as airlines, and trains. It could provide access to additional data sources and insights. So collaborating with travel industry partners will provide us the opportunities for real-world testing and validation of the prediction model and recommendation engine.

References

[1] B. E. Moscow, *Tourism in bangladesh*, https://bangladeshembassy.ru/about-bangladesh/tourism/, [Accessed: May 9, 2023] (cit. on p. 1).

[2] ShareTrip, *Sharetrip - the largest travel portal in bangladesh*, http://www.sharetrip.net/, [Accessed: May 9, 2023] (cit. on p. 1).

[3] Bangladesh Tourism Board, *Official website of bangladesh tourism board*, http://tourismboard.gov.bd/, [Accessed: May 9, 2023] (cit. on pp. 1, 3).

[4] B. E. Vienna, *Tourism in bangladesh*, https://bangladootvienna.gov.bd/tourism-in-bangladesh/, [Accessed: May 9, 2023] (cit. on p. 1).

[5] TripAdvisor, *Bangladesh tourism: Best of bangladesh*, https://www.tripadvisor.com/Tourism-g293935-Bangladesh-Vacations.html, [Accessed: May 9, 2023] (cit. on p. 1).

[6] Gozayaan, *Gozayaan - travel to bangladesh and beyond*, https://www.gozayaan.com/, [Accessed: May 9, 2023] (cit. on p. 2).

[7] Obokash, *Obokash - your travel partner*, https://www.obokash.com/, [Accessed: Month Day, Year] (cit. on p. 2).

[8] Dragon Holidays BD, *Dragon holidays bd - travel agency in bangladesh*, https://dragonholidaysbd.com/, [Accessed: May 9, 2023] (cit. on p. 10).

[9] Amy Bangladesh, *Amy bangladesh - the largest online marketplace in bangladesh*, https://www.amybd.com/index?v=987654484, [Accessed: May 9, 2023] (cit. on p. 11).

[10] Nijhoom Tours, *Best places to visit in bangladesh*, https://nijhoom.com/best-places-to-visit-in-bangladesh/, [Accessed: May 9, 2023] (cit. on p. 14).

[11] Business Haunt, *Top 10 places to visit in bangladesh*, https://businesshaunt.com/top-10-places-to-visit-in-bangladesh/, [Accessed: May 9, 2023] (cit. on p. 14).

[12] T. Ghani, N. Jahan, S. H. Ridoy, A. T. Khan, S. Khan and M. M. Khan, 'Amar bangladesh - a machine learning based smart tourist guidance system,' in *2018 2nd International Conference on Electronics, Materials Engineering Nano-Technology (IEMENTech)*, 2018, pp. 1–5. DOI: 10.1109/IEMENTECH.2018.8465377 (cit. on p. 15).

[13] V. Parikh, M. Keskar, D. Dharia and P. Gotmare, 'A tourist place recommendation and recognition system,' in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2018, pp. 218–222. DOI: 10.1109/ICICCT.2018.8473077 (cit. on p. 15).

[14] M. E. B. H. Kbaier, H. Masri and S. Krichen, 'A personalized hybrid tourism recommender system,' in *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, 2017, pp. 244–250. DOI: 10.1109/AICCSA.2017.12 (cit. on p. 15).

[15] I. Patil and P. Navangul, 'Tourist place recommendation system using machine learning,' 2021 (cit. on p. 15).

[16] P. A. Manjare, M. P. V. Ninawe and M. M. L. Dabhire, 'Recommendation system based on tourist attraction,' *Imperial journal of interdisciplinary research*, vol. 2, 2016 (cit. on p. 15).

[17] E. Pantano, C.-V. Priporas and N. Stylos, ''you will like it!' using open data to predict tourists' response to a tourist attraction,' *Tourism Management*, vol. 60, pp. 430–438, 2017, ISSN: 0261-5177. DOI: https://doi.org/10.1016/j.tourman.2016.12.020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0261517716302680 (cit. on p. 16).

[18] R. Jiang and H. Jiang, 'Personalized cruise travel recommendation system based on data mining and glonass tools,' in *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2022, pp. 1172–1175. DOI: 10.1109/ICSSIT53264.2022.9716354 (cit. on p. 16).

[19] R. Logesh and V. Subramaniyaswamy, 'Exploring hybrid recommender systems for personalized travel applications,' in *Cognitive Informatics and Soft Computing*, P. K. Mallick, V. E. Balas, A. K. Bhoi and A. F. Zobaa, Eds., Singapore: Springer Singapore, 2019, pp. 535–544, ISBN: 978-981-13-0617-4 (cit. on p. 16).

[20] P. Nitu, J. Coelho and P. Madiraju, 'Improvising personalized travel recommendation system with recency effects,' *Big Data Mining and Analytics*, vol. 4, no. 3, pp. 139–154, 2021. DOI: 10.26599/BDMA.2020.9020026 (cit. on p. 16).