GRADUATE PROGRAMS INSTITUTE

ANALYZING THE KEY FACTORS AFFECTING TRAVEL RECOMMENDATION SYSTEM BASED ON RANDOM FOREST APPROACH

Muhammad Zia UL HAQ

MASTER'S THESIS

Istanbul, November 2024

ANALYZING THE KEY FACTORS AFFECTING TRAVEL RECOMMENDATION SYSTEM BASED ON RANDOM FOREST APPROACH

Muhammad Zia UL HAQ

MASTER THESIS

COMPUTER ENGINEERING DEPARTMENT
COMPUTER ENGINEERING MASTER'S PROGRAM WITH THESIS

MASTER'S THESIS ADVISOR:

Asst. Prof. Dr. Gizem TEMELCAN ERGENECOŞAR

MASTER'S THESIS JURY MEMBERS:

Prof. Dr. Abdurazzag Ali A ABURAS

Asst. Prof. Dr. Zeynep GÜVEN AYDIN

PREFACE

In the culmination of my academic journey, I am delighted to present this master's thesis entitled "Analyzing The Key Factors Affecting Travel Recommendation System Based On Random Forest Approach".

This thesis represents the culmination of months of research, analysis, and reflection, and it is with a sense of accomplishment and enthusiasm that I share my findings with the academic community. I extend my deepest gratitude to my thesis advisor, **Asst. Prof. Dr. Gizem Temelcan Ergenecosar,** whose guidance and expertise were invaluable throughout this process. Their insightful feedback, constructive criticism, and unwavering support significantly enriched the quality of this work. Without her I could not do anything. I am also thankful for the assistance and collaboration of **Asst. Prof. Dr. Ozlem Feyza Erkan**, whose contributions enhanced the multidimensional nature of this research.

Additionally, I am grateful to my friends and family for their unwavering encouragement and understanding during this demanding period of study.Their support provided the emotional sustenance necessary for navigating the challenges of research and writing. May this work serve as a stepping stone for further inquiry and foster a spirit of intellectual curiosity.

CONTENTS

ABSTRACT

ANALYZING THE KEY FACTORS AFFECTING TRAVEL RECOMMENDATION SYSTEM BASED ON RANDOM FOREST APPROACH

This thesis focuses on the development and evaluation of a travel recommendation system. The objective is to analyze the factors effecting the travel recommendation system and propose a recommendation system for the travelers in Pakistan. By leveraging advanced algorithms and user preference data, the system aims to enhance user satisfaction and streamline travel planning. The system provides personalized travel suggestions by processing user preferences using the Random Forest algorithm. The research addresses the challenges of making recommendations with limited data availability and explores how data expansion can enhance accuracy. For the role of algorithmic efficiency and data quality on system performance in travel recommendation systems, the key components of these systems are analyzed. Recommendation accuracy, user satisfaction, and overall system performance from the basis for evaluating the specified system. The thesis suggests practical implementation in real-world scenarios and provides insights into potential system improvements. Due to limited available data, the system requires further dataset expansion for more accurate recommendations.

Keywords: Travel Recommendation System, Random Forest, User Satisfaction, Travel Planning, Recommendation Algorithms.

Date: 01/11/2024

ÖZET

RASGELE ORMAN YAKLAŞIMINA DAYALI SEYAHAT ÖNERİ SİSTEMİNİ ETKİLEYEN TEMEL FAKTÖRLERİN ANALİZİ

Bu tez, bir seyahat öneri sisteminin geliştirilmesi ve değerlendirilmesine odaklanmaktadır. Amaç, seyahat öneri sistemini etkileyen faktörleri analiz etmek ve Pakistan'daki gezginler için bir öneri sistemi sunmaktır. Gelişmiş algoritmalar ve kullanıcı tercih verilerini kullanarak, sistem kullanıcı memnuniyetini artırmayı ve seyahat planlamasını kolaylaştırmayı amaçlamaktadır. Sistem, kullanıcı tercihlerini rasgele orman (random forest) algoritması kullanıp işleyerek kişiselleştirilmiş seyahat önerileri sunmaktadır. Bu çalışma, sınırlı veri erişilebilirliği ile önerilerde bulunmanın zorluklarını da ele almakta ve veri setini genişletmenin doğruluğu nasıl artırabileceğini göstermektedir. Seyahat öneri sistemlerinde algoritmik verimlilik ve veri kalitesinin sistem performansı üzerindeki rolü için, bu sistemlerin ana bileşenleri analiz edilmektedir. Tavsiye doğruluğu, kullanıcı memnuniyeti ve genel sistem performansı, belirtilen sistemi değerlendirmek için temel oluşturmaktadır. Tez, pratik bir uygulama önermekte ve potansiyel sistem iyileştirmeleri hakkında bilgiler sunmaktadır. Sınırlı veri seti nedeniyle, sistemin daha doğru önerilerde bulunabilmesi için veri setinin daha da genişletilmesi gerektiği sonucuna varılmaktadır.

Anahtar Kelimeler: Seyahat Öneri Sistemi, Rasgele Orman, Kullanıcı Memnuniyeti, Seyahat Planlaması, Öneri Algoritmaları.

Tarih: 01/11/2024

ABBREVATIONS

GPS : Global potioning system

GLONAS : Globalnaya Navigazionnaya Sputnikovaya Sistema

MAP : Mean Average Precision

MAE : Mean Absolute Error

POI : Point of Interest

PLS-SEM : Partial Least Structural Equation Modeling

RMSE : Root Mean Square Error

SVM : Support Vector Machine

UAT : User Acceptance Testing

WDI : World Development Indicator

LIST OF FIGURES

LIST OF TABLES

YÖK watermark

1. INTRODUCTION

With the introduction of the latest technologies, the travel sector has made significant growth. In this modern era, the travel recommendation plays a vital role in different aspects with the introduction of new tools for the modern day's travelers. The most important one is to navigate the huge amount of information available in order to make precise decisions for the travelers in this technological era. These systems generate individualized travel recommendations by utilizing complicated algorithms, data analysis, and content generated by users. However, a number of important aspects affect how accurate and reliable these systems are. This thesis aims to analyze key aspects affecting travel recommendation systems, including customer trust and transparency, infrastructure and technological constraints, limited local content and information, financial constraints and payment methods, and language and cultural barriers. The following titles provide explanations of these key aspects.

1.1. Customer Trust and Transparency

Mohammadi et al. (2019) delve into one of the essential component of a recommendation system customer trust. The dependency and accuracy of the system is the foundation to built trust with the accurate information provided to the users. Mohammadi et al. (2019) furthermore shed light on users reviews and rating they see. The reviews and ratings has to be totally genuine and unbiased. Different strategies are used for building trust, like detailed reviewer profile , verified reviews and how the system provides suggestions to the users. Roy et al. (2022) open communication about data collection, storage, and usage practices guarantee clients of their privacy and security. Professional and timely customer service is also critical for building and maintaining trust, as it ensures that user problems and issues are addressed quickly and effectively.

1.2. Infrastructure and Technological Constraints

Chaudhari et al. (2019) evaluate the effectiveness of travel recommendation systems is heavily dependent on the quality of the underlying infrastructure and technology. High-speed internet access and reliable mobile connectivity are crucial for users to access these platforms seamlessly. In many parts of the world, especially in developing regions, limited technological infrastructure poses significant challenges. Users may experience slow loading times, frequent disconnections, and difficulty accessing the platform altogether. Cunha et al. (2020) elaborate the performance and scalability of the recommendation system itself are critical. The platform must be able to handle high traffic volumes, integrate with other services such as booking platforms, maps, and local transport systems, and provide a seamless user experience. Technological advancements such as cloud computing, artificial intelligence, and machine learning can help overcome some of these constraints, but they require significant investment and technical expertise.

1.3. Limited Local Content and Information

Chaudhari et al. (2020) present a comprehensive survey of travel recommender systems, and discusses the difficulties that arise from the relatively underexplored or undeveloped locations that can offer limited and outdated data compared to main tourist attractions. The quality and the accuracy of local information have a significant impact on the effectiveness of travel recommender systems. Whereas regularly visited overseas destinations offer a wealth of information, less frequented or faraway regions may lack thorough and up-to-date information. This can pose a major obstacle for tourists looking for unique or less-traveled adventures. Xiang et al. (2021) discussed the authentic local information that can only be collected through direct collaboration with local businesses, tourism agencies, and content creators with practical experience and in-depth knowledge of the area. In order to make sure that user-generated content is served in a way that can genuinely help to close these gaps, it does firstly need to be selected and verified above all else its consistency. Which are indeed not easy to store a vast and varied database over different sites with valid and up-to-date information. Brabham et al. (2013) provide information about the crowdsourcing platforms which are community-driven and enabled by technology have the potential to enhance local content, however checking the quality through quality control and moderation is necessary if such platforms were to work at their best.

1.4. Environmental Factors

Building a good travel recommendation system in Pakistan requires the integration into its development of environmental and seasonal aspects. According to Hidaka et al. (2020), the research focuses on most of the current recommendation systems target providing recommendations on tourist sites and desirable routes prior the actual travel and are not dynamic enough to adapt the recommendations to the current state of affairs, including national or local weather conditions, traffic jams, and closed facilities during the travel. Sui and Lin (2020) in response comes up with new strategies that use real-time data collected on site and that are therefore more accurate and individualized. One of them, outlined in, employs mobile technology to collect situational awareness information such as weather, opening hours of attractions, and other relevant information and subsequently employ artificial intelligence methods to recommend the most appropriate means of transport for tourists.

1.5. Financial Constraints and Payment Methods

One of the key factors that potentially influences the usage of travel recommendation system is financial constraint. Asaithambi et al. (2023) provide significat insights into the users belonging to different financial backgrounds when it comes to using the platform and services of the system. The last but equally important elements are payment system security and accessibility. Furthermore, there are differences in various parts of the world in payment methods, some people use credit cards, digital wallets and debit cards while others prefer to pay with cash. Therefore, in order to offer an efficient payment process, additional issues such as exchange rates, transaction costs, and protection for payments must be resolved. Moreover, Xiang et al. (2020) conducted study based on the business models of travel recommendation systems, namely paid placement of content, affiliate marketing, and paid membership models, may affect the level of objectivity and credibility of the recommendations made. Possessing customers' data generate revenue and yet protecting customers' trust and serving them with impartial recommendations.

1.6. Language and Cultural Barriers

Travel recommendation systems face significant challenges due to different culture and language. O'Hagan et al. (2013) discuss the importance of accurate translation and cultural sensitivity, which is applicable to travel recommendation systems. Multinational support is indispensable in travel recommendation systems, but poorly implemented translation can lead to a negative experience. For example, a mistranslation might lead to misunderstanding or not following the preffered recommendation, destroying the user experience. On the other hand, the travel recommendation systems should be sensitive to the cultural norms of different languages. Ricci et al. (2015) explored the relevant context for the discussion on optimizing recommendations based on cultural and regional differences. For instance, not all languages use the same norms, realities or restaurants or cafes, and other facilities. These systems must confront a number of significant challenges in order to reach their full potential and satisfy users. Step essential including trust building among customers with transparency of your services and robust data privacy measures making sure to overcome infrastructure and technological limitations by enriching local content, accommodating financial constraints and diverse payment methods and bridging language and cultural gaps.

This thesis aims to examine these aspects, providing a comprehensive analysis of the current state and future directions for travel recommendation systems. By addressing these challenges, travel recommendation systems can enhance their reliability, user satisfaction, and overall effectiveness, ultimately transforming the travel experience for users in Pakistan.

1.7. Problem Statement

With the rise of online travel platforms and recommendation engines, people have been able to plan their trips in a personalized way, based on user preferences. Nevertheless, ensuring transparency, especially user trust of these systems remains a huge challenge. While concerns about its validity, fairness and transparency persist, especially in Pakistan, where technological constraints, poor local content and cultural barriers add additional hurdles. This thesis creates an analytical framework for analyzing travel recommendation systems, which are specifically designed to address challenges that travel in Pakistan presents. Maintaining user trust and transparency, overcoming infrastructure and technology limitations, as well as being held accountable to financial and payment methods, environmental factors and language and cultural barriers are key challenges. Despite these constraints, the framework will look at factors such as user satisfaction, data privacy, access, scale, and accuracy of the recommendations in order to determine system efficiency. The analysis will also investigate methods to improve transparency in recommendation algorithms, increase the reliability of user reviews, and work with other data sources to balance the absence of local content. The aim is to offer insights and advice, in the process of enhancing the travel recommendation system to better accommodate the needs and expectations of its users as well as maintaining trust and transparency in the traveller's decision making journey in Pakistan.

Figure 1.1 shows the relationship between trust and transparency, infrastructure and technological constraints, environmental factors, limited local content and information, financial constraints and payment methods as well as cultural and language barriers in a travel recommendation system.

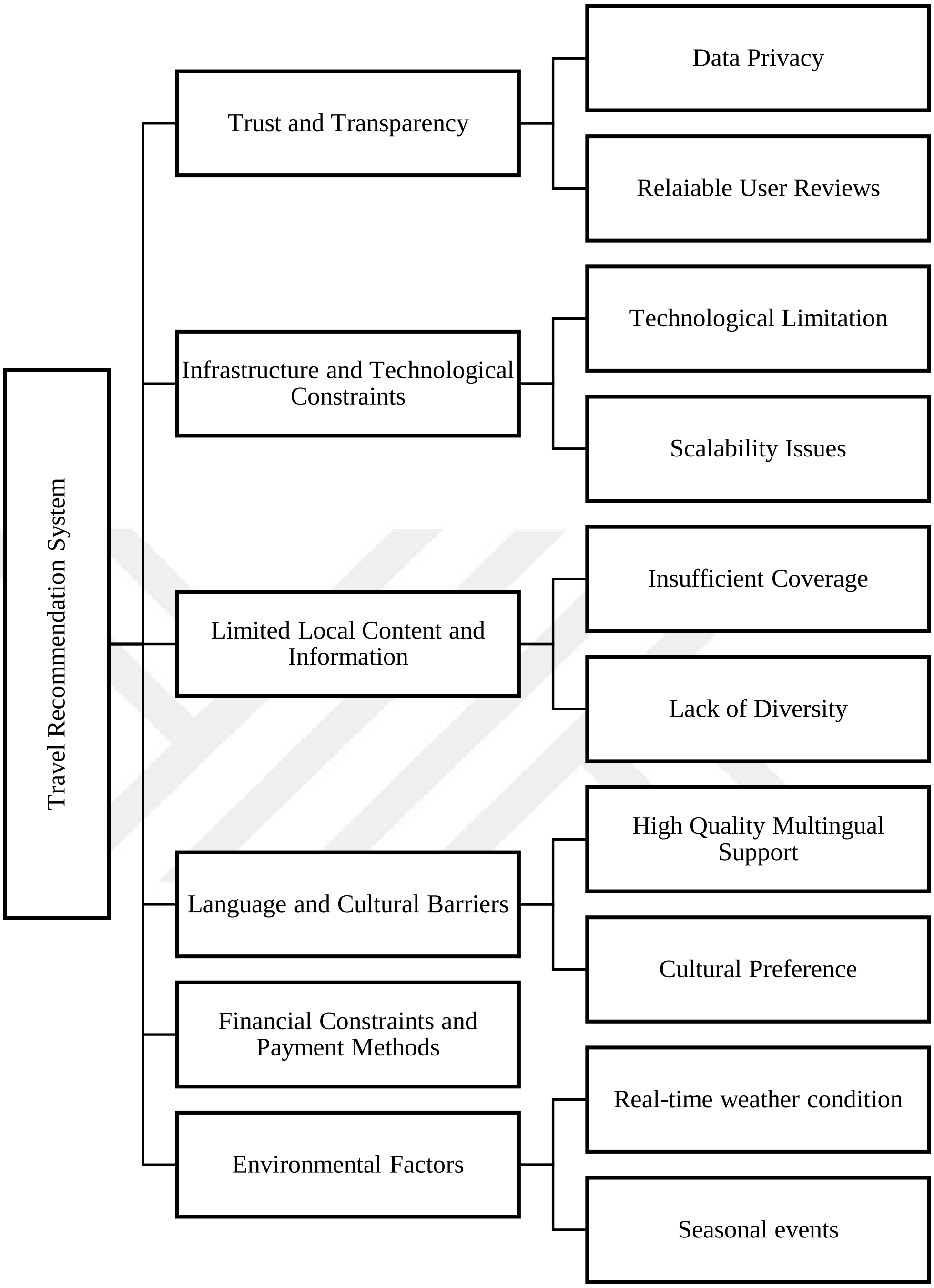

Figure 1. 1 Factors effecting travel recommendation system

In this diagram:

- The central component is the travel recommendation system.

- The main challenges outlined in the problem statement: “customer trust and transparency, infrastructure & technological constraints, limited local content, environmental factors, payment methods and cultural and language barriers."
- Each challenge is further expanded to illustrate specific aspects.
- The arrows indicate the interaction and interdependence between the challenges and the recommendation system.

This diagram provides a visual representation of factors effecting travel recommendation system, emphasizing the interconnected nature of the challenges and the need to address them comprehensively within the travel recommendation system.

1.8. Motivation

In today's modern era, the travel recommendation systems have become essential tool for travelers seeking personalized and reliable travel advice. However, several challenges hinder the optimal functioning of these systems, including issues related to customer trust and transparency, infrastructure and technological constraints, and the availability of local content, language and cultural barriers, environmental factors and as well the payment methods. The objectives are addressing these challenges which are crucial for several reasons:

Enhancing User Trust and Experience: Trust is a cornerstone of any recommendation system. Travelers need to trust that the recommendations are unbiased, accurate, and tailored to their preferences. Enhancing transparency in how recommendations are generated can significantly improve user satisfaction and trust, leading to higher user engagement and loyalty.

Improved Decision Making: Transparent recommendation processes empower users with insights into how recommendations are generated, what data is used, and how algorithms operate. This transparency enables users to trust the recommendations provided and make better-informed travel decisions.

Increased User Engagement: Trust and transparency foster trust between users and platform providers, encouraging greater engagement with the platform, increased interaction with recommendations, and higher conversion rates for bookings and reservations.

Ethical Considerations: Addressing concerns related to data privacy, algorithmic bias, and ethical considerations is essential for promoting responsible and ethical use of user data within recommendation systems. Transparent practices build user trust and confidence in the platform's ethical standards.

Industry Competitiveness: Travel companies that prioritize trust and transparency in their recommendation systems can gain a competitive edge by differentiating themselves from competitors and attracting and retaining loyal users.

Overcoming Technological Barriers: Infrastructure and technological limitations can restrict the accessibility and efficiency of travel recommendation systems. By identifying and mitigating these constraints, the system can become more robust and accessible to a broader audience, including those in areas with limited technological infrastructure.

Expanding Local Content Coverage: Accurate and comprehensive local information is vital for making relevant travel recommendations. Addressing the scarcity of local content can improve the quality and relevance of the suggestions provided by the system, making it more useful for travelers exploring less-documented destinations.

Driving Innovation and Improvement: Analyzing and addressing these challenges can drive innovation in the development of travel recommendation systems. This can lead to the creation of more advanced algorithms, better data integration techniques, and overall improvements in system performance.

Supporting Sustainable Tourism: By providing reliable and transparent travel recommendations, these systems can promote sustainable tourism practices. They can guide travelers to lesser-known destinations, distribute tourist traffic more evenly, and support local businesses by highlighting unique local attractions and services.

1.9. Goal of the Thesis

A travel recommendation systems that focuses on customer trust, technological constraints, local content limitations, environmental factors, cultural and language barrier payments methods and environmental factors are essential for improving these systems. This will not only enhance user satisfaction and trust but also contribute to more accessible, innovative, and sustainable travel experiences. The goal is to propose a recommendation system which will sort out the issues faced by the the customers related to the factors already discussed above in Pakistan . The proposed system figure out the issue related to Pakistan while using the travel recommendations system. There are different aspects which this thesis look for in order to propose a comprehensive and reliable recommendation system which can be trusted by the customers and also easy to use for the travelers in Pakistan.

2. LITERATURE REVIEW

Travel recommendation systems have witnessed a significant surge in popularity and utility in recent years, primarily due to the introduction of online travel platforms and the increasing volume of available travel information. These systems leverage advanced technologies such as data analytics, machine learning algorithms to analyze user preferences and historical data, thereby generating personalized recommendations for various travel related entities such as destinations, accommodations, activities, and transportation options. The success of travel recommendation systems does not only rely on the accuracy and relevance of their recommendations but also on the trust and confidence that users place in these systems. Users must feel assured that the recommendations provided are credible, reliable, and aligned with their preferences and interests. Moreover, the transparency of the recommendation process is essential for users to understand how recommendations are generated and to assess the system's credibility and impartiality.

The reliable and trusted system gains the confidence of its users which enhances the customer's pleasure and engagement with it. On the contrary, lack of confidence in the system stop users from using and accepting it. Trust in recommender systems Gavalas et al. (2014) actually boast about the ways in which recommender systems have the potential to build user trust through transparency of the recommendations thus improving the systems' efficiency as well as enhancing user satisfaction.

Among the available literature, a work by Massa and Avesani. (2007) is comprehensive enough. The authors also discuss trust and distrust in recommender systems, where they touch on the issue of transparency as the possible path to the elimination of distrust among users. They emphasize on how openness of the recommendation models will improve the satisfaction of the users and their interactions.

Transparency as a key concept in creating confidence in information systems Glushko et al. (2000) express that transparency is of big significance for increasing confidence in information systems, and one should always make an attempt to explain the processes occurring in an information system in simple words necessary for understanding. His work applied to travel recommendation systems where the use of transparency increases the confidence of the users in the provided recommendations.

Paskahlia Gunawan and Tho (2021) explores the expansion and advancemen made due to the incorporation of the latest technologies in the travel business. Consumers express their perception of a specific place for tourism, accommodations, and packages on the various sites accessible over the internets.

Kim and Kim (2017) examined the influence of customer reviews on travel recommendation system. Customers strongly depend on online reviews, which causes skepticism concerning the honesty and openness of the recommendation source based on the findings of various surveys done, it shown that customer opinions and reviews play a major role on the recommendation system. The appropriate adopted measures were made with regard to the context of the present study. All the items were on 5- likert type scale which include; strongly disagree (1), disagree (2), neutral (3), agree (4) and strongly agree (5).

There are some essential specifications concerning consumer trust and transparency issues explored by Joo et al. (2022) in their work to explains accuracy of the system. Due to the centrality of the accuracy of a travel recommendation system in issues of customer trust and absolute disclosure, it fulfills traditional objectives. To ensure that customers get the right recommendations for their travel needs, the efficiency of the system used is of measurable value to the customers. The credibility of a travel recommendation system is defined as the perceived accuracy of the system as well as the extent to which the system can be believed to deliver accurate results. To conduct the study, the researchers used a meta-analysis methodology. The meta-analysis methodology allows the researchers to synthesize and analyze previous quantitative findings from multiple empirical studies.

Customers have to be convinced that the recommendations given by a system are sounding sensible and made on the basis of accurate information. Paskahlia Gunawan and Tho (2021) paper sufficiently describes two approaches to analyze the customer needs which are content based filtering and collaborative filtering. From such insights, it can be argued that travel recommendation systems would benefit from the inclusion of transparency to build trust from customers. Such features which include, the accuracy of the products produced, the credibility of the results generated, and the transparency of the entire process all work towards fostering customer confidence in a system.

A review of guidelines for transparent and explainable AI systems for a travel recommendation context by coman et al. (2022). The paper outlines principles of transparent explainable and accountable AI for a travel recommendation systems. They point out that transparency is the key aspect within acceptance and trusting of the AI-supported systems stressing that whatever should be easily explained to the users.

Reviewing user-technology trust in the Context free recommendations Hsieh et al. (2023) focuses on the transparancy of recommendation systems. Therefore, according to the presented study, it highlights the importance of information openness when it comes to the recommendation formation in order to enhance the user's confidence and satisfaction provided services within the hospitality sector. Hybrid recommendation has been adopted by the authors to enhance the accuracy of the recommendation. The results of the best performance has been generated using machine learning techniques in order to discover what the customers prefer.

In the current day context it has been noted that the travel industry has gone through great improvements with the help of the new technologies and the abundant information available in web. Thus, the travel recommendation system has been really helpful to the users in choosing the destination they wanted, where to stay, and food to eat among other interests.

Law et al. (2014) conducted a survey to get the users' reviews and point of views, providing essential awareness about the users preferences and behaviors. The questions posed to the users in the survey are based on satisfaction, factors that are relevant to the trust in the specific system and customers' perceptions about the reliability of the system.

Kuang et al. (2019) one specific study was carried out to systematically evaluate the literature on travel recommendation system, and many valuable themes were found. Some of these themes were, user-generated content, the importance of algorithms and data sources, the importance of customer's satisfaction on travel recommendation system . The research found out that in order to gain the customers satisfaction, it is compulsory the consumers must trust the system. The weighted matrix factorization method was aaassused in this paper because it surpasses other forms of factorization models in terms of POI suggestion, but also that recommendation performance is enhanced when the spatial clustering phenomena is considered.

Joo et al. (2022) investigated user preferences and behavior analysis for travel recommendation and this study underscored the importance of context-aware recommendations tailored to individual user characteristics, shedding light on factors influencing recommendation effectiveness. The study investigate user preferences and behavior analysis for travel recommendation, stressing the need for context-aware recommendations tailored to individual characteristics.

Xu et al. (2020) focusing on location recommendation techniques, this work discusses strategies to leverage user location data for more relevant travel suggestions, recognizing the significance of geographical context in enhancing recommendation quality. The study discusses location recommendation techniques, emphasizing strategies to leverage user location data for more relevant suggestions.

Yin et al. (2015) proposing a collaborative filtering-based approach to tourist attraction recommendation, this study utilizes user-item interaction data to refine recommendations, demonstrating the effectiveness of collaborative filtering in enhancing recommendation accuracy. A collaborative filtering-based approach to tourist attraction recommendation, refining recommendations using user item interaction data.

Introducing a tourist recommendation system integrating user preferences and geographical location. Gao et al. (2015) focused on identifying factors affecting the engagement of users to recommendation systems. Their research also provided insight into decision features which make the use of products better by users including customized settings, design of the user interface and the openness of algorithms used to offer recommendations. Some recommendations for deeper and more efficient work of such systems were given during this work with focus on the analyses of user interactions and feedback.

Advocating for a hybrid collaborative filtering approach for recommendation Hsieh et al. (2023) work incorporates geo-social information to enhance recommendation accuracy, demonstrating the effectiveness of integrating multiple data sources in recommendation systems.

Providing insights into trust and distrust in recommender systems Massa et al. (2007) study emphasizes the importance of transparency in fostering user trust, highlighting the need for transparent recommendation processes to enhance user satisfaction. It offers insights into trust and distrust in recommender systems, highlighting the importance of transparency. Trust matric algorithm is used in order to enhance the the capability of the system whos main objective was to do prediction based on the trust network, the trustworthiness of unknown users.

Discussing principles for building transparent AI systems applicable to travel recommendation systems Coman et al. (2022) study emphasizes the importance of transparency in fostering user trust, highlighting the need for transparent recommendation processes to enhance user satisfaction. It highlights the need to give clear explanations of the approaches adopted by the expert.

Exploring the impact of social influence on travel recommendation acceptance Pan et al. (2020) research study emphasize the role of social connections in shaping user attitudes and behaviors towards recommended destinations and activities. It examine the impact of social influence on travel recommendation acceptance, emphasizing the role of social connections. Different surveys and online website are used to collect the data and analyze the effects of social influences.

Proposing a hybrid recommender system for tourist attractions Jannach et al. (2014) work combines content-based and collaborative filtering approaches to provide more diverse and personalized recommendations, catering to different user preferences and interests. The authors proposed a hybrid recommender system for tourist attractions, combining content-based and collaborative filtering approaches.

Introducing a context-aware recommendation model for travel planning Gao et al. (2015) study considers factors such as user preferences and trip constraints to generate more relevant and personalized recommendations, enhancing the overall user experience. introduce a context-aware recommendation model for travel planning, considering user preferences and trip constraints. Leveraging user-generated content for hotel recommendation, this research enhances recommendation relevance by incorporating user reviews and ratings, providing users with more informative and trustworthy recommendations.

Travel recommendations were reviewed in a systematic literature which is discussed by Lian et al. (2014). The study delve into dozen themes highlighting how travelers select their points of interest before embarking on travelling. The topics discussed were, an extent to which the customer trust affects the trip recommendation systems, transparency in the system algorithms and data resources and ways how user-generated content contributes to trust and transparency. As a result of this research, it has been established that the clients' trust is of great significance in travel recommendation systems utilization and acceptance. Customers are more likely to depend upon and apply recommendations from systems which they believe are trusted.

Developing a trust-aware recommendation model for travel services Wang et al. (2018) develop a trust-aware recommendation model for travel services, considering user preferences and trust relationships. The study considers both user preferences and trust relationships to generate more reliable and trustworthy recommendations, fostering user trust and confidence in the system.

Muneer et al. (2022) explained the integration of sentiment analysis into travel recommendation systems in their study. Their work was however more focused on improving the quality of the recommendations based on the sentiment and mood of the users. When users opt for their choices from a feeling towards it this makes he suggestions more closer since the system leads his users to the options he or she has a liking.

Wang et al. (2019) proposed a novel recommendation model integrating temporal dynamics for travel planning, capturing evolving user preferences. The study captures evolving user preferences over time, providing users with more adaptive and personalized recommendations based on their changing interests and preferences.

Discussing the impact of mobile recommender systems on travel decision-making, Gavalas et al. (2014) emphasizes the importance of real-time recommendations in facilitating spontaneous travel planning and decision-making, enhancing the overall travel experience. The paper aimed at discussing the flow of recommendation systems, focusing on the service characteristics of some of them debating about the open problems in the related field. Therefore, the review procedure is systematic, according to classification system that refers to the method of observation from three distinct points of view in the assessment of existing tourism recommendation systems. Their chosen architecture engage users in the delivery regarding the set of recommendations and the parameters considered for arriving at recommendations.

User engagement and satisfaction in travel recommendation system Flicker et al. (2024) explores the fundamental precondition for the acceptance of digital services, with transparency being a key mechanism to enhance trustworthiness. However, there are contradictions and complexities in how transparency affects trust. For instance, while transparency is generally seen as positive, an overwhelming complexity of information can impede trust. This suggests that the relationship between transparency and trust is dependent on individual consumer perceptions and literacy.

Xu et al. (2020) study of mobile-based short video apps for travel information highlights the importance of perceived enjoyment and usefulness, which are crucial for fostering trust in the technology. Similarly, Adiwijaya et al. (2015) discuss the role of vendor trustworthiness, including ability, benevolence, and integrity, is emphasized as influential in customer trust and purchase intentions in online environments.

Harn and Raheem. (2023) discuss the need for a system arises from the rapid growth of the tourism industry and the necessity for more effective tools to help travelers make informed decisions. Collaborative filtering approach based on user similarities is used by the authors. The collaborative filtering technique does not require domain knowledge since embeddings are automatically learnt, and the matrix factorization can solve the sparsity problem. The proposed matrix factorization model achieved an average MAP (Mean Average Precision) value of 0.83, with RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) values being 1.36 and 1.24, respectively. These metrics indicate that the model was able to generate personalized location recommendations effectively based on users' past visits and interactions with other users.

Xu et al. (2020) research has highlighted technological affordances and the perceived ease of use and usefulness of platforms to enhance trust. The challenge remains in balancing transparency with the complexity of information to avoid overwhelming users.

One significant constraint is the stage of tourism is information technology application Zhang et al. (2009) is in a state of ofness, particularly in rapidly developing tourism scenarios . This can limit the functionality and responsiveness of travel recommendation system. Additionally, the precision of positioning technologies, such as the Global Positioning System , can be a limiting factor, as seen in the case of cruise travel recommendation systems that require accurate location data. Jiang et al. (2015) has addressed the integration of different systems, such as the Globalnaya Navigazionnaya Sputnikovaya Sistema (GLONASS), is one approach to overcoming these limitations.

Nitu et al. (2021) have shed light on the challenges of effectively exploiting the relationship between users and their points of interest is magnified by the need to consider user preferences, which are dynamic and can change over time . This necessitates the development of advanced algorithms capable of adapting to these changes, such as those incorporating time-sensitive recency weight.

Paskahlia Gunawan and Tho (2021) article reviewed the issues of adopting travel recommendation systems in third-world countries. Their study identified two major obstacles, such as structural and technological barriers, and content and information scarcities. To overcome these challenges, they proposed to use existing technologies that are currently in use in the modern society, these include the mobile gadgets and internet connection to acquire and disseminate travel information. The gathering of proper authentic data concerning the tourist destinations should involve the local governments, travel agencies as well as other stakeholders. With the help of surveys and other techniques for collecting user data as well as obtaining validated data from other sources, it is possible to build a couple of powerful data pools.

Shen et al. (2016) argued that the accuracy of recommendations may suffer due to the low frequency of tourism and the diversity of attraction styles across different cities, which are not always adequately captured by existing systems. Diverse collective intelligence is propose to do better personalized recommendation for the users. The authors have contributed to enhance user engagement with travel recommendation systems and have emphasized the importance of understanding users' mental models. The paper has explained about the interactive nature of humans and that is the reason different web based applications are there for the users so that they can get benefit out of it by choosing their own preferences of choosing a destination or moving from one place to another.

Werthner et al. (2006) suggests the concept of Ambient Intelligence for that future travel recommendation systems which could become more adaptive and personalized by learning from user behavior in real-time. Machine learning approach is used which includes the K-NN, decision tree and SVM (support vector machine) to enhance the recommendation process for the users of the system.

Zhang et al. (2009) discuss the design and functionality of travel recommendation systems, the researchers do not explicitly address financial constraints or payment methods within their frameworks. To obtain the capacity of item-based model better extendable ability than the ontology-based model approach, the authors have used the proposed ontology-based user's item vector space representation of user model.

Chaitra et al. (2022) does mention the use of Google APIs, which could potentially include payment services, but the paper does not investigate the specifics of handling financial constraints or payment methods within the travel recommendation system. Furthermore, the study offers general overview of functionality of these systems, and the potential for integrating payment services is implied . There is a lack of focused analysis on financial aspects within the context of travel recommendations. Further research would be required to explore this dimension in depth.

Simuka et al. (2024) while not directly related to travel recommendation systems, does explore the use of technology in university canteen systems, including the implementation of cashless payment systems. This indicates a broader trend towards integrating financial transactions and user preferences into service-oriented platforms, which could be applicable to travel recommendation systems as well.

Paskahlia Gunawan and Tho (2021) give a comprehensive analysis of local content and information which play a crucial role in travel recommendation systems, as they provide valuable insights and suggestions for tourists looking to explore a particular destination. The application leverages the Dijkstra algorithm to calculate the shortest path for visiting various tourist attractions, providing an optimized itinerary for travelers. The paper mainly focuses on Dijkstra algorithm to provide the shortest route and also the most reliable route for the travelers from one destination to anaother.

Due to the relatively fewer advancements of modern technologies in third world countries Heeks et al. (2018) talked about a possibility of having very few local contents and information on these recommendation systems. This can be as a result of numerous factors including lack of capital, structures as well as dismal internet access in these countries. Therefore, the tourists may encounter alot of difficulties in getting the relevant, current and comprehensive information about the tourist places, available accommodations, transport facilities, and cultural and social practices. In this sense, the systems may be severely limited by the lack of local content and information about third world countries in particular.

The efficiency of travel recommendation systems is severely limited by technological limitations, especially in underdeveloped nations. The lack of digital resources, inadequate internet connectivity, and technical limitations that many third-world nations experience make it challenging for recommendation systems to compile enough local content. Travel networks are so frequently skewed toward more developed areas with superior technology infrastructure. Heeks et al. (2018) emphasizes how financial limitations and restricted access to contemporary communication technologies impede the growth and use of ICT systems, particularly travel platforms, in low-income nations. Furthermore, the World Bank highlights the significance of tackling the "digital divide" in order to improve access to tourism-related information in developing nations.

Also, some travel platforms are using technology which is not able to process large datasets quickly and it relies slowly and in the meantime, without integration with machine learning. Similarly, Xiang et al. (2020) note that small tourism destinations would not be granted a change in their travel advisories owing to financial constraints. This will potentially make them less competitive.

The other challenge, which stands in the way of proper services utilization, is the use of old technologies. Smooth travel experience for the passengers with multiple safe payment options but payment option restrictions also help customers in many places to forget travel suggestion systems entirely. Zhou et al. (2019) has a significant problem because the international gateways have no connectivity with local facilities and there are many foreign tourists which prefer several payment gateways. It brings out the severe issues for travelers and travel agents that certain nations have cash dependency, no digital wallets, and unsafe online transactional frameworks as the writers point of interest.

The travel industry is experiencing a period of drastic change, as technological innovations, novel financial structures, and evolving social perspectives merge to transform how travel suggestions are formulated. As Xiong et al. (2015) explained, technical boundaries such as current infrastructure restrictions and the need to seamlessly interweave disparate platforms present formidable hurdles for travel businesses aiming to deliver customized, streamlined recommendations.

Moreover, Hou et al. (2022) developing simplified transaction models across global markets and reconciling linguistic and cultural variances aggravates the design and application of these algorithms, as discussed. Simultaneously, while some foresee reaping rewards from big data analytics, privacy issues and ethical complexities loom large, demanding circumspect solutions from all stakeholders in the emerging digital nomad economy.

Hou et al. (2022) explains the necessity for navigational advisors has expanded in equal measure. However, these present pathways struggle to synchronize suggested itineraries with momentary milieus, contextual clues, and individual tastes in real-time.

Outdated technologies are another limitation. There must be appropriate secure payment methods and additional features for enhancing the experience of passengers. Ratan et al. (2012) state that after surfacing the recommendation systems, a limitation of practice was encountered whereby payment alternatives are not available for people in many jurisdictions. Moreover, the authors explain such a gap brings a limitation where there is no interaction between international payment platforms with local systems since most foreigners tend to pay differently. There is not only an impact on the travelers but on travel agencies as well. In other words, they are pretty much losing out because of a possible cash addiction of the host countries, a lack of online payment options for local currencies, or a secure space for making online payments.

As stated by Brabham et al. (2013) crowdsourcing can increase quality of material localization. However, user-generated content may contain errors or may not be useful for the specific language preference with no quality control and translation. Based on Paskahlia Gunawan and Tho (2016), one potential way of boosting of user satisfaction and relevancy across the regions could be to include the language subtitles for travel recommendation platforms.

Note that cultural differences may lead to a large effect on the efficiency of the travel advice systems. Ricci et al. (2015) pointed out that cultural preferences vary from one country to another. It is difficult to create a versatile recommendation system that will satisfy different consumers.

Zhang et al. (2020) also examine cultural and social factors that effect users’ perception and interaction with the travel recommendation systems. Any suggestion that may be considered culturally insensitive, for instance recommending non-halal foods to Muslim tourists, is likely to make the concerned parties dissatisfied and have little confidence in the system. The authors stress the need for developing flexible solutions that can provide suggestions about culturally relevant aspects of a certain country, including rituals, traditions, and constraints regarding the diet. For instance a tip that is good for the visitor using a website in the Western Europe region may not be ideal for a user in Asia or the Middle East region due to differences in lifestyle and expectations. To eliminate or reduce these problems several methods that have been put forward.

In their paper, Brabham et al. (2013) suggest improvement in crowdsourcing and AI-based quality control to offer culturally sensitive material count. Integration of machine learning and human decision-making enhances the reliability and ethnicity of the suggestions.

While O'Hagan et al. (2013) propose a more comprehensive approach to language translation, it remains apparent that Artificial Intelligence translations should be employed to improve on the works of translators and ensure the imperative cultural nuances are well captured. Thus, following the studies of Zhou et al. (2019) it is valuable for global payment providers to cooperate with local financial institutions to extend the range of payment solutions and upgrade the user experience across different markets.

Ali Raza et al. (2023) investigate the relationships between three environmental triggers, i.e., awareness, concern, and knowledge, and tourists' environmental attachment and green motivation, which influence the pro environmental behavior within Pakistan's tourism sector. The data were gathered from 237 domestic tourists with structured questionnaire and then analyzed by using the SmartPLS, the results indicate all three triggers (habitual usage, imagery usage, comparison usage) are positive to environmental attachment and green motivation, promoting pro-environmental behavior. Furthermore, the study investigates the moderating role of moral obligation, which moderates the relationship between ecological attachment to the environment and pro environmental behaviour. This research provides insight for improvement on sustainable tourism practices in Pakistan. They have used closed ended questions with a five point scale to gather responses valid by adopting the items of an existing study. This study applied Partial Least Structural Equation Modeling , squares which was used to analyze data and examine both the measurement and structural models.

Waqas el at. (2018) aim of the study is to assess the impacts of the multiple environmental factors on the region's tourism sector in Pakistan. To achieve this purpose, there are three primary dimensions of the ambient environment, which are economic and financial indicators, social development, and macroclimatic changes, with their requisite indicators, through the data set of the world development indicator (WDI). On the basis of significant statistical tests and techniques, separate standardized regression equations are developed for the three factors. The study findings indicate that (a) economic and financial indicators have great impact on private and (b) social development and macro climate indicators have positive and negative influence on the tourism industry of Pakistan. Statistical significance of all the models and also provided strong evidence for the key managers to perform strategic planning for the tourism industry when considering the stated factors of the study.

Nisa. (2024) Environmental impacts of tourism on local communities in Pakistan, particularly in Hunza and Diamer districts is explored in the paper. It highlights the provision of economic growth, by tourism, and addresses the emerging environmental and socio-economic challenges facing local communities. The study highlights the necessity of sustainable tourism practices to be utilized by inducing minimum negative impacts, while improving the community's resilience to it. Secondary qualitative research is used in the study and the available. Results of the research show that tourism has had a very negative impact, the most notable negative impacts being that deforestation, biodiversity loss, pollution, and socio economic inequalities. The results imply that decision and service makers should adopt sustainable tourism principles to protect the environment and preserve local cultures.

3. RESEARCH METHODOLOGY

This chapter outlines the detailed process used to create the tourist destination recommendation system. It starts by explaining the research design, which is essentially the plan that guided the entire thesis. The research design helps to ensure that every step taken is purposeful and aligned with the project’s goals. Furthermore, the discussion on data acquisition is given, which refers to the methods used to gather the information needed to build the system. This data might include details about various tourist destinations, user preferences, and more. The quality and relevance of this data are crucial for the system's effectiveness. Following this, it examines the analytical techniques. These techniques are methods used to process and understand the data. They help to identify important patterns and trends that the system can use to make accurate recommendations. It also covers the machine learning models implemented in the thesis. Machine learning involves training computers to learn from data so that they can make predictions or decisions without being explicitly programmed to do so. These models are the core of the recommendation system, enabling it to suggest destinations that match user preferences. Finally, the chapter places special emphasis on the formulas and methodologies used throughout the thesis, ensuring that every step is thoroughly explained and understood. This comprehensive approach ensures that the recommendation system is both accurate and reliable.

3.1 Overview of the Research

The research starts with the objective of evaluating information on tourist attractions in Pakistan. The purpose of this thesis is to suggest tourist destinations based on the input provided by the users. The research plan involves selection of machine learning algorithms, identification of of the features related to the dataset which are extracted and following the sequencial process to refine the models for obtaining better accuracy or result. The collected dataset consist of various types of tourism spots in different regions of Pakistan. The dataset is sourced from kaggle, provided by (Salman, 2024). Kaggle holds a large variety of datasets provided by the community and which are used primarily for analyzing, training models, and other purposes associated with data management. It is consist of diverse set of information to train the model. Before using the machine learning algorithms, the data is preprocessed and relevant features is extracted and establishing iterative procedure for better travel recommendation. The findings include final model's accuracy in suggesting most suitable destination for the users.

3.2 Research Design

The research design is quantitative, focusing on machine learning techniques to suggest tourist destinations. Figure 3.1 the steps for research design is illustrated in flowchart.

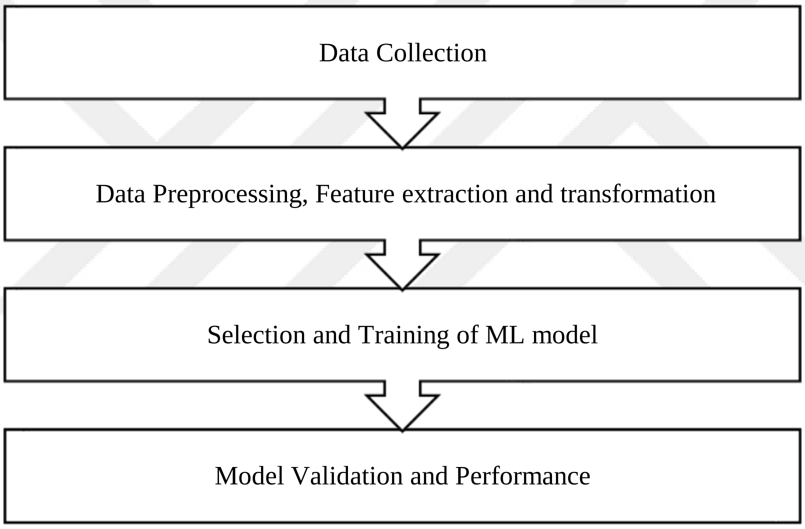

Figure 3. 1 Steps of quantitative research

3.3 Data Acquisition

Data acquisition refers to the process of collecting the information we need to train the recommendation system. In this case, we needed data on tourist destinations in Pakistan, such as information about popular sites, user reviews, ratings, and possibly demographic information of users like age, interests, or travel history. This data is crucial because of the quality and relevance of the information. The data gathered directly impact how well our recommendation system will work. The sourced data is from an online platform. The data is then transformed to remove any inconsistencies or irrelevant information.

3.3.1 Datasets

The dataset used in this study is acquired from a platform known for its extensive collection of datasets used for machine learning and data science projects. Kaggle is known for its extensive repository of datasets, ranging from beginner friendly to highly specialized collections. It provides a space for data scientists, researchers, and developers to access data, share their projects, and participate in competitions. For this thesis, it is selected as the source of our dataset because of its reliability and the quality of datasets it hosts, particularly those related to machine learning and data analysis tasks. The dataset includes seventy rows and six columns. Places are classified according to their types such as, lakes, moutains, hill stations, islands, waterfalls etc. It is consist of different columns which provides essential information about every place. It provides the nature of each place whether it is a tranquil lake, sacred lake or mountanous area. At the same time it provides detail information about the destination, it's longitude, latitude to provide the exact location. Also, it provides brief information about the nearby areas of the recommended location. The dataset is based on the provinces of Pakistan, namely Punjab, Khayber pakhtunkhwa, Gilgit-Baltistan, Sindh, Azad Kashmir and Balochistan. The collection covers a variety of sites, ranging from the charming beauty of Ansoo Lake with its distinctive tear shaped form to the historical value of Lahore's Badshahi Mosque. Mountains such as Baltoro Glacier and Concordia are well-known for being part of the Karakoram Range, which contains some of the world's highest peaks. This gather data represents a wide spectrum of activities Pakistan has to offer visitors, from mountain adventure to lake peacefulness and cultural exploration in historical cities. It's an invaluable resource for anyone wishing to experience the country's rich scenery and legacy, and it's crucial to your thesis, particularly for predicting places in a trip recommendation algorithm. The key attributes includes of the dataset include a following.

Destination descriptions: Textual descriptions provide detailed information about each location. The first key attribute of the dataset is the textual descriptions of each destination. These descriptions provide detailed information about the tourist spots, offering insights into what makes each location unique. For example, a description might include the historical significance of a place, the natural beauty of a site, or the cultural relevance of a landmark. These descriptions are crucial for the recommendation system because they help to capture the qualitative aspects of each destination. By analyzing this text, the system can identify key phrases and keywords that are indicative of user preferences. For instance, if a user expresses interest in historical sites, the system can search through the descriptions for terms like ancient, heritage, or ruins to find relevant destinations.

Geographical coordinates: Latitude and longitude values indicating the precise location of each destination. Another vital attribute is the geographical coordinates, which include the latitude and longitude of each destination. These coordinates provide the exact location of a tourist spot, allowing the system to understand the spatial distribution of the destinations. Geographical data is particularly useful when incorporating location-based filters in the recommendation system. For example, if a user wants to explore destinations within a certain radius from their current location, the system can use the coordinates to calculate distances and suggest nearby options.

Categories: Classification of destinations into types such as lakes, mosques, islands, and mountainous regions. These categories help to group similar destinations together, which is useful for both the recommendation system and the end user. For the recommendation system, categorization allows for more targeted suggestions. If a user is specifically interested in exploring natural landscapes, the system can prioritize destinations categorized as lakes, mountains, or islands. Similarly, users interested in cultural or religious tourism might prefer destinations categorized under mosques or historical sites. The categorization also helps in refining the recommendation algorithms. By understanding the commonalities between different categories, the system can make better predictions about what types of destinations a user might enjoy.

Districts: Indicating the administrative region of each destination. The dataset includes information about the administrative districts in which each destination is located. This is another important attribute because it provides a geographical context that is more granular than just the country level. Knowing the district allows the recommendation system to offer suggestions based on administrative boundaries, which can be useful for users planning to travel within specific regions. For example, a user planning a trip to the northern areas of Pakistan might be more interested in destinations located in districts like Gilgit-Baltistan or Khyber Pakhtunkhwa. The district data also helps in understanding regional tourism patterns. For instance, the system can analyze which districts are more popular among tourists or have more attractions, thus fine-tuning its recommendations based on regional popularity. The dataset is selected based on its comprehensive nature, offering rich contextual and geographical information essential for developing a robust recommendation system.

The selection of this dataset was not random but was based on its comprehensive nature and the rich cntextual and geographical information it provides. Each attribute in the dataset plays a critical role in ensuring that the recommendation system can deliver accurate and relevant suggestions to users. Why this dataset was chosen and how each attribute contributes to the robustness of the system is discussed in deatail in the following.

3.4 Techniques for Analyzing the Data

3.4.1 Data Analysis

The data analysis is conducted using both descriptive and predictive techniques. These approaches are explained in detail in the following.

Descriptive analysis: Provided insights into the distribution and characteristics of the data, helping to inform the feature selection process. This technique is like looking at a picture and describing what you see. Examine the data to understand its basic features and patterns. For example, we might look at the most popular tourist destinations, the average distance between them, or the types of places people like to visit (e.g., beaches, mountains, cities). This information helps to decide which parts of the data are most important for building a recommendation system.

Predictive analysis: Involved the application of machine learning algorithms to forecast tourist destinations based on user preferences. This is like trying to predict the future. We use machine learning algorithms to guess which tourist destinations a person might like based on their preferences. For example, if someone likes beaches, the system might suggest other beach destinations. Or, if someone has visited a specific city and enjoyed it, the system might recommend similar cities. This technique helps to provide personalized recommendations to users.

3.4.2 Implementation of the Analytical Techniques

The implementation involved the use of Python libraries such as Pandas for data manipulation, Scikit-learn for machine learning, and regular expressions for text processing. The following key steps were involved:

3.4.3 Data Loading and Preprocessing:

Data loading and preprocessing are essential steps in any data analysis or machine learning project. Some aspects include the use of Pandas library for code implementation, loading of data from a CSV file and creating DataFrame based on this data. After which, this DataFrame is fit for data processing, analysis or even data visualization. It provides the representation of your data in Python form where the data is mostly readable and well formatted for doing any action on the given data. These steps involve getting your data into a format to work with and clean enough to use in your models.

Thid code part in Figure 3.2 provided Python code uses the pandas library to load data from a CSV file. It starts by importing the pandas library, which is a powerful tool for data manipulation and analysis in Python, and assigns it the alias "pd" for easier use in the script. The pd.read_csv function is then used to read a csv file. This function loads the data into a pandas DataFrame, a two-dimensional table-like data structure that allows easy handling of the data. The resulting DataFrame is assigned to the variable df, which can then be used for various operations, such as analysis, filtering, or visualization of the information related to tourist destinations. This step is essential for efficiently managing large datasets and preparing the data for further processing or analysis.

```
Import pandas as pd
df= pd.read_csv('tourist_destinations.csv')
```

Figure 3. 2 Loading dataset from csv file

Feature Extraction: A custom function is used to extract features from the Description column. The written descriptions are extracted from description column of the DataFrame df and only the numerical altitude values are used by the code. After that, a new column with the name altitude is generated, to accommodate these data. If an altitude is not provided within the description, then the matching entry in the altitude column will be None. This makes it possible to capture and analyse altitude information from complex text descriptions in a systematic way.

Figure 3.3 shows the extraction of features from the description column of dataset.

```
import re
def extract_features(description):
altitude = re.search(r'altitude of (\d+)', description) return altitude.group(1)
 if altitude else None df['altitude'] =
df['Desc'].apply(extract_features)
```

Figure 3. 3 Extraction of features

Label Encoding: The code transforms the categorical data in the category and district columns of the DataFrame df into numerical values using label encoding. This is a necessary step for many machine learning algorithms that require numeric input. The original string categories are replaced with integers, and the corresponding LabelEncoder objects are stored in the label_encoders dictionary for potential future use.

The Figure 3.4 presents the transformation of data present in the category section of the dataset.

```
from sklearn.preprocessing
import LabelEncoder label_encoders = {} for column in ['category', 'district']:
le = LabelEncoder() df[column] = le.fit_transform(df[column])
label_encoders[column] = le
```

Figure 3. 4 Transformation of categorical data

3.5 Machine Learning Technique

The machine learning techniques employed to predict tourist destinations based on user input are explained. The approach focuses on the selection of an appropriate model, training procedures, validation techniques, and performance evaluation metrics.

The core task of recommending future tourist trends involves predicting the type of tourist destination that aligns with user preferences. For this task, the Random Forest Classifier is used**.** The Random Forest algorithm is selected due to its robust performance in classification tasks and its ability to handle both numerical and categorical data. Random Forest Classifier ensemble learning method operates by constructing multiple decision trees during training. The Random Forest does not rely on a single tree for making predictions. Rather, it compiles forecasts from every tree in the forest. The Random Forest takes the average of all the trees' predictions when it comes to regression tasks, which include making continuous value predictions. When the objective involves categorization, the model selects the most frequently predicted answer as the final output by holding a vote. The Random Forest decreases the possibility of overfitting, which occurs when a model learns the training data too well and performs badly on fresh, unseen data.

3.5.1 Training the Model

The training process begin with splitting the dataset into training and testing subsets. The used method is train test split, from Scikit-learn, which randomly divides the dataset while maintaining a typical 70-30 split between training and testing data.

3.5.2 Data Preparation

Input Features: The features selected for training the model included category, district, latitude, and longitude. These features are used to train the model. Category tells us what type of place that is whether it is a mountain, lake, or historical site. It is stored as text and therefore treated as a string. This is another text, or a string, deciding the district, the which region or city the place is in i.e. Islamabad or Lahore. These two things have to be reduced to numbers for the model to take into account. Latitude and longitude are numbers that tell you exactly where something is located on the map. It tells us where the place is and after what whether it is north or south and how far east or west it is. As floating point numbers they are stored. When combined all of these features together help the model learn patterns based on not only the type of destination, but also the location.

Target Variable: The target variable for the suggestion task is the category of the tourist destination.

Model Training: The Random Forest Classifier was then trained on the training data which are the input features and targeted variables. The model was configured with 100 trees (n_estimators=100), which is a typical choice for balancing performance and computation time.

The training process involves the model learning patterns from the input features that correlate with the target variable. The random state parameter ensures that the results are reproducible.

3.5.3 Model Validation Techniques

Validation is a critical step to ensure that the model generalizes well to unseen data. The following techniques are employed for validation:

Train-Test Split Validation: The accuracy of the model was initially evaluated on the test set. This basic validation method provides an indication of how well the model performs on data it has not seen during training.

Cross-Validation: To further validate the model's robustness, cross-validation is employed. This technique involves splitting the dataset into *k* folds and training the model *k* times, each time using a different fold as the test set and the remaining *k-1* folds as the training set. The cross-validation score is the average accuracy across all k folds, providing a more reliable estimate of model performance.

3.5.4 Accuracy

Accuracy measures the proportion of correctly predicted instances out of the total instances in the test set. In this thesis, the Random Forest model achieved an accuracy of approximately 79.19% on the test set. This indicates that the model correctly predicts the type of tourist destination for about three-quarters of the test cases.

3.5.5 Confusion Matrix and Classification Report

Beyond accuracy, the examined the confusion matrix and the classification report, which provide more detailed insights into the model's performance across different categories.

Confusion Matrix: It provides a detailed description of what the model did, this includes the number of correct predictions, wrong predictions and the type of error made. Normally, the matrix is arranged in a way of square matrix, where the rows stand for the real class, which is True label, while the columns stand for the predicted class Class from the model. The confusion matrix gives a breakdown of true positives, true negatives, false positives, and false negatives, helping to identify which categories the model predicts well and where it may be making mistakes.

The confusion matrix is it not only gives the number of correct and wrong predictions, but also informs which specific classes are causing problems for the model. This data is crucial for iterative model's improvement, as it shows what went wrong, when model is better at one class prediction than the other, and whether it is generally sensitive to, for instance, false positive or false negative errors. From the confusion matrix you can decide how you want to alter the model, for instance the class weights can be changed, the hyperparameters could be tuned or more data can be collected for the specific classes. The model performs well on several classes (e.g., classes 3, 8), but there are a few misclassifications (such as class 4 being predicted as 3). Overall, the confusion matrix suggests that the model is making mostly correct predictions but still has room for improvement in correctly distinguishing some classes.

The following given Figure 3.5 shows the number of correct and worong predictions based on the available data made by confusion matrix.

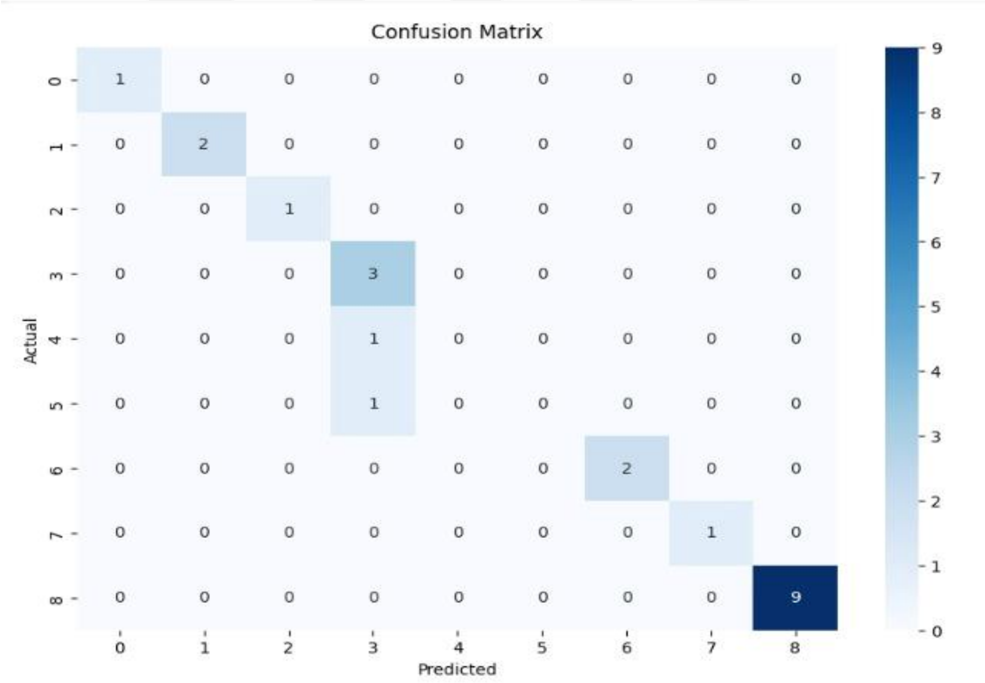

Figure 3. 5 Confusion matrix prediction graph

Classification Report: The classification report includes precision, recall, and F1-score for each category, offering a comprehensive view of the model's strengths and weaknesses. The class report offers a detailed comparison of the model, and it perform about the strengths and weakness with the help of the precision, recall, and F1-score for each class. The categorization report shows the performance of your model over the number of classes you have defined, and it offers a quantitative characterization of the general performance of the model. It is very helpful when handling cases where some data points are more numerous than others or when a problem under consideration is a multi-class classification problem. By looking at the accuracy of the model, as well as the precision, recall and F1-Score for each class you can establish in which area the model is weak and in which area it is strong. By using the confusion matrix not only it is possible to find out the total accuracy, but also decide how well a classifier is doing per class, whether is more inclined towards false positives or false negatives and where changes should be made either to the data, or the model being used. Several of the performance factors reported herein are used in conjunction to enable a cross-comparison among them. This means that depending on the demand you have put into the search engine you will be in a position to either select a model that has a better recall value or the one that has a higher precision. Table 3.1 below shows the classification report.

Table 3. 1 Classification report

Class label	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	1
2	1.00	1.00	1.00	2
3	1.00	1.00	1.00	1
5	1.00	0.00	0.75	3
9	0.60	0.00	0.00	1
10	0.00	1.00	0.00	1
11	0.00	1.00	1.00	3
14	1.00	1.00	1.00	1
15	1.00	1.00	1.00	9
Accuracy			0.90	21
Macro Average	0.73	0.78	0.75	21
Weighted Average	0.85	0.90	0.87	21

3.6 Random Forest Model

The trained Random Forest model is used to predict the category of tourist destinations based on user input. Users provide inputs related to their preferred district, latitude, and longitude. The model uses these inputs to predict the most likely category of tourist destination. The category is then matched with the dataset to retrieve the corresponding tourist destination's details, providing the user with a personalized recommendation.

3.7 Relationship between User Preferences and Model Recommendation

Understanding the relationship between user preferences and model predictions was crucial. By analyzing how different input variables influenced the predictions, the model was refined to better match user expectations. The use of the Random Forest model allowed for an in-depth analysis of feature importance, helping to fine-tune the model's predictive capabilities. Creating an efficient recommendation model requires an understanding of how user preferences and model suggestions relate to one another. In this procedure, many input variables like user interests, habits, and demographic data are extensively examined to see how they affect the model's predictions. It remains as a useful process through which one can gain some insights of why a defined model comes up with certain decisions and to what extent these decisions meet the expected expectations of the users.

An efficient technique for determining which input variables have the greatest influence on the model's predictions is feature importance analysis. You can figure out which features of the data are most crucial for accurate estimates by weighing the significance of each one. This realization enables the model to be adjusted, maybe by adding or subtracting unnecessary features or by stressing specific ones more or less, improving the fit between the model's outputs and user expectations.

By means of iterative analysis and improvement, the model gains greater sensitivity to user preferences, hence augmenting its capacity to produce precise and pertinent predictions. This improves the model's performance and guarantees that the results it produces are more in line with what consumers are really searching for.

4. FINDINGS, RESULTS AND DISCUSSION

This chapter outlines the findings that have been achieved on the subject of the development and evaluation of the travel recommendation system for travelers in Pakistan. It evaluates the efficiency of the system, influence of constraints and effectiveness of the recommendation provided by the system. Also, which factors have an impact on the effectiveness of the system: user satisfaction, availability of data, and the performance of the algorithm and, how to further improve the results' accuracy and the speed of calculations.

4.1 Summary of Findings

This work's major contribution is the development of this system, which has an overall accuracy rate of 79.19% fairly high in light of the dataset restrictions. This suggests that for about three quarters of test samples the system was able to top up and suggest appropriate tourist destinations that fit the users' preferences. However, the performance differs by the type of tourist destinations that are focused on. Some types are recommended more accurately than the others, for instance, mountain destinations and lakes while some areas such as historical sites were not well suggested due to lack of local data. This implies that even though the system can deliver, the efficiency it delivers depends on the size and variability of the data set.

With gradio, machine learning users take advantage of an interactive web interface to their models. The benefit of the approach is that you have the ability to input data, receive predictions, and immediately see results. Widgets for district and category are dropdown menus and a number of nights is a numeric field. The tourist point, description and approximate budget will be shown as text displayed in the output. Integrating gradio is easy, and it comes with an intuitive interface that doesn't require you to write anything fancy with frontend code. Users can interact with the machine learning model and immediately get feedback. Inputs for the recommendation system can be customized for specific purpose. Figure 4.1 shows the interface used for destination recommender system.

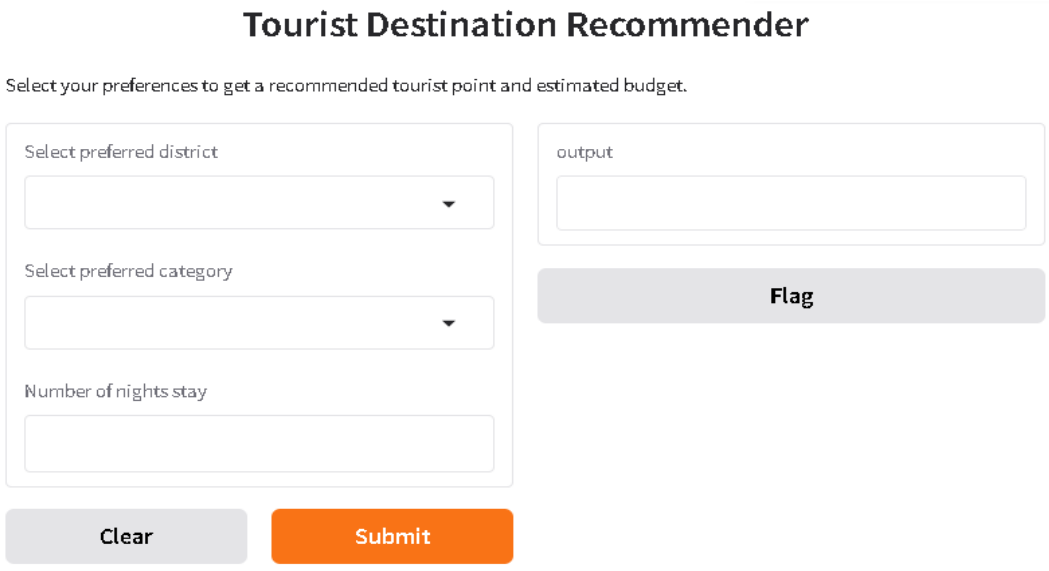

Figure 4. 1 Destination recommender system interface

Gradio interface is designed to capture user input for three key parameters. It also includes district, category, and the number of nights to stay. The user selects district, category and nights of stay. A machine learning model is used to process this information thus indicating a recommended tourist spot and a cost estimate. Based on the user's input it presents the recommendation in the output interface of your gradio application. After the user makes their selections, It will show the tourist destination according to the chosen category and district. An extracted portion of the recommended place is given, and the extracted portion is referred to as database identifier. The estimated cost is provided to the user based on the entries user makes. The implementation of this interface provides an organized and clear way to display feedback, which all while using practical travel tips to improve user experience.

Figure 4.2 shows the recommended destinations based on the user's selections from the input entries.

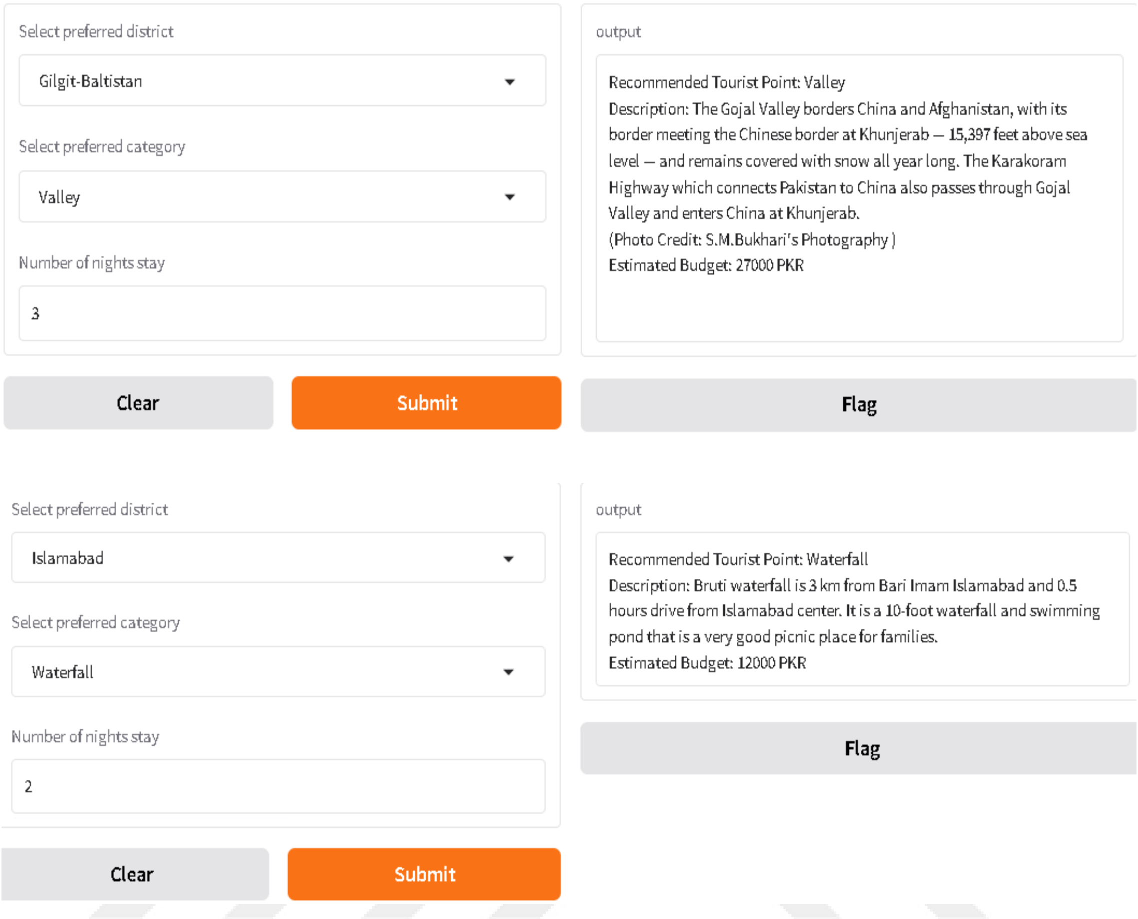

Figure 4. 2 Input and Output of the system interface

4.2 Technology and Data Constraints

The major challenge faced during this research is the lack of recent local data. The data set containing information about tourist attractions in Pakistan was obtained from Kaggle and though the data was rather limited, it ensured that the system was not able to make fair and balanced recommendation. However, the Random Forest algorithm that is implemented to the system worked well, providing relatively good results given the limits of the data. Moreover, the state of the technological and informational environment in Pakistan affected the effectiveness of implementing the system particularly in the field that is characterized by a lack of widespread use of digital information. Even during the preparation of this paper, the system failed to give the right recommendations of distant places because there are no reviews online and real-time information.

4.3 Recommendations Accuracy and User Satisfaction

Upon the assessment, it became clear that the existing correlations between the parameters meant a high dependence of overall user satisfaction on the relevance of the recommendations provided. The algorithm proved to efficiently recommend the most popular and described tourist attractions with a good percentage of success. At the same time, however, for the niche and the less popular destinations, especially for the places that typically have less popular travel search and have less-popular travel searches and lowest proportion of visitors and travelers, the system is found to be performing relatively worse and wanted to be improved with the call for using broader data and integrating real-time data sources for increasing the breadth of the recommendations and for increasing the depth and accuracy of recommendations. Users also want a list of recommendations filtered by preferences as well as recommendations that are linked to current temperature or some other variable. Ideal management of such dynamic factors would have enhanced user satisfaction as well as the utility of the system in question.

4.4 Cultural and Language Barriers

The research also suggests the requirement for a new approach to address language and culture in travel recommendation systems. Since the dataset was mostly English and for an international audience, there are other cultural and language correlative oversights which were not well captured. For example, suggestions like religious places, cultural events, etc are less favorable as there is no localized content available, and the quality of translation is not very good in most cases.

4.5 Payment Methods and Financial Constraint

Second, aspects of finance and modes of payment in finance exert severe influence on the evaluation of the travel recommendation system. Recommendations is way out of budget for users in Pakistan and in some cases payment options were simply foreign to a lot of users. However, making payment options location based, and also by taking into consideration economic factors, will make the system more useful to a larger audience.

4.6 Future Directions for Improvements

With the introduction of advance technologies there has been significant growth in the usage of the travel recommendation system. Customers want to visit their dream destination with respect their preferences and choices. The main focus of this thesis is based on the hurdles faced by the users in Pakistan. The reasons users find it difficult is to use these platform to plan their trip. The proposed system focuses on issues faced by the customers related to travel recommedation systems in Pakistan.The data is collected to provide a sytem that can sort out these issues with respect to to customers trust and transparency , infrastructure and limited local informationc, environmental factors, cultural and laguage barriers and technological constraints and payment methods for the users in Pakistan.

However, there's left room for improvement in the different services of the propose system, and the system needs huge amount of investment in order to launch in different parts of the country. Precisely, when we target the third world countries like Pakistan, it's necessary for the people to use the system in their mother tongue, so, the propose system will be available in Urdu in future for the users. Furthermore, there is a need of getting big amount of data by doing local surveys or from the available platforms. Additionally, the system has to be made more user friendly that would be easy for every user to get benefit out of it. More work should be done for the personalized recommendation to provide more accurate services for the customers. Building reputation is an important factor for a new system because it's difficult for the users to trust a new system in the market, that's why there's a need to make system reach out for the users. For all these mentioned services budget is the key factor with the high computational requirement high cost is also needs to be paid.

The current system provides a sound foundation from which travel recommendations in Pakistan can be made, but more changes are required before it reaches its optimal. These include data expansion. This compilation contains more elaborate data, that is, users' contributions, reviews, and real-time information acquired locally. The system will further be enhanced in order to ask more about cultural and language preference which will make it more culturally and language relevant to the local users. It attempts to make the whole recommendation process more transparent so users can trust how recommender works and what factors are considered.

REFERENCES

Asaithambi, S. P. R., Venkatraman, R., and Venkatraman, S. (2023). A thematic travel recommendation system using an augmented big data analytical model. Technologies, 11(1), 28.

Adiwijaya, M. (2015). The Effect of Vendor Trustworthiness toward Online Purchase Intention through Costumer Trust.

Avgerou, C. (2010). Discourses on ICT and development. Information Technologies and International Development, 6(3), 1–18. https://doi.org/10.1162/itid.2010.00089

Brabham, D. C. (2013). Crowdsourcing: A model for leveraging online communities. In A. Delwiche & J. J. Henderson (Eds.), The Participatory Cultures Handbook. Routledge.

Brabham, D. C. (2013). Crowdsourcing: A model for leveraging online communities. In A. Delwiche & J. J. Henderson (Eds.), The Participatory Cultures Handbook (pp. 120-129). Routledge.

Chaudhari, K., and Thakkar, A. (2020). A comprehensive survey on travel recommender systems. Archives of Computational Methods in Engineering, 27(4), 1545–1571. https://doi.org/10.1007/s11831-019-09363-7

Cunha, G., Peixoto, H., Machado, J. (2020). Improving Performance of Recommendation System Architecture. In: Analide, C., Novais, P., Camacho, D., Yin, H. (eds) Intelligent Data Engineering and Automated Learning – IDEAL 2020. Lecture Notes in Computer Science(), vol 12490. Springer, Cham. https://doi.org/10.1007/978-3-030-62365-4_47.

Chaudhari, K., and Thakkar, A. (2020). A comprehensive survey on travel recommender systems. Archives of Computational Methods in Engineering, 27(4), 1545–1571. https://doi.org/10.1007/s11831-019-09363-7

Chaitra, D., Prasad, V. R. B., and Vinay, B. N. (2022). A Comprehensive Travel Recommendation System. Smart Innovation, Systems and Technologies, 248. https://doi.org/10.1007/978-981-16-4177-0_62

Coman, T. (2022). CIRSE 2022 Book of Abstracts. CardioVascular and Interventional Radiology, 45(S4), 145–1035. https://doi.org/10.1007/s00270-022-03246-4

Flicker, K., Rauber, A., Kern, B., and Ekaputra, F. J. (2024). Factors Influencing Perceptions of Trust in Data Infrastructures. International Journal of Digital Curation, 18(1). https://doi.org/10.2218/ijdc.v18i1.921

Gao, H., Tang, J., Hu, X., and Liu, H. (2015). Content-Aware Point of Interest Recommendation on Location-Based Social Networks. http://foursquare.com

Gavalas, D., Konstantopoulos, C., Mastakas, K., and Pantziou, G. (2014). Mobile recommender systems in tourism. In Journal of Network and Computer Applications (Vol. 39, Issue 1, pp. 319–333). https://doi.org/10.1016/j.jnca.2013.04.006

Heeks, R. (2018). Information and Communication Technology for Development (ICT4D). Routledge. https://doi.org/10.4324/9781315652657

Harn, C. W. S., & Raheem, M. (2023). Recommendation system on travel destination based on geotagged data. International Journal of Advanced Computer Science and Applications, 14(5),117–128. https://doi.org/10.14569/IJACSA.2023.0140511

Hsieh, Y. C., Lu, L. C., & Ku, Y. F. (2023). Review evaluation for hotel recommendation. Electronics (Switzerland), 12(22). https://doi.org/10.3390/electronics12224673

Harn, C. W. S., and Raheem, M. (2023). Recommendation System on Travel Destination based on Geotagged Data. International Journal of Advanced Computer Science and Applications, 14(5), 117–128. https://doi.org/10.14569/IJACSA.2023.0140511

Hsieh, Y. C., Lu, L. C., & Ku, Y. F. (2023). Review Evaluation for Hotel Recommendation. Electronics (Switzerland), 12(22). https://doi.org/10.3390/electronics12224673

Jannach, D., Zanker, M., and Fuchs, M. (2014). Leveraging multi-criteria customer feedback for satisfaction analysis and improved recommendations. Information Technology and Tourism, 14(2). https://doi.org/10.1007/s40558-014-0010-z

Jiang, S., Qian, X., Shen, J., Fu, Y., and Mei, T. (2015). Author topic model-based collaborative filtering for personalized POI recommendations. IEEE Transactions on Multimedia, 17(6). https://doi.org/10.1109/TMM.2015.2417506

Joo, K. J., Bae, S. H., Pyo, J. W., Ghosh, A., Park, H. J., and Kuc, T. Y. (2022). Autonomous Navigation with Active SLAM for Disinfecting Robot. International Conference on Control, Automation and Systems, 2022-November, 112–116. https://doi.org/10.23919/ICCAS55662.2022.10003703

Kim, S. B., and Kim, D. Y. (2017). Antecedents of corporate reputation in the hotel industry: The moderating role of transparency. Sustainability (Switzerland), 9(6). https://doi.org/10.3390/su9060951

Kuang, H., Chang, L., Bin, C., and Gu, T. (2019). Review of a context-aware travel recommendation system. CAAI Transactions on Intelligent Systems, 14(4). https://doi.org/10.11992/tis.201901013

Law, R., Buhalis, D., and Cobanoglu, C. (2014). Progress on information and communication technologies in hospitality and tourism. International Journal of Contemporary Hospitality Management, 26(5), 727–750. https://doi.org/10.1108/IJCHM-08-2013-0367

Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., & Rui, Y. (2014). GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 831–840. https://doi.org/10.1145/2623330.2623638

Mohammadi, V., Rahmani, A. M., Darwesh, A. M., and Sahafi, A. (2019). Trust-based recommendation systems in Internet of Things: a systematic literature review. Human-centric Computing and Information Sciences, 9, 1-61.

Mohammadi, V., Rahmani, A.M., Darwesh, A.M. (2019). Trust-based recommendation systems in Internet of Things: a systematic literature review. Hum. Cent. Comput. Inf. Sci. 9- 21. https://doi.org/10.1186/s13673-019-0183-8.

Muneer, M., Rasheed, U., Khalid, S., and Ahmad, M. (2022). Tour Spot Recommendation System via Content-Based Filtering. 2022 16th International Conference on Open Source Systems and Technologies, ICOSST 2022 - Proceedings. https://doi.org/10.1109/ICOSST57195.2022.10016820

Nitu, P., Coelho, J., and Madiraju, P. (2021). Improvising personalized travel recommendation system with recency effects. Big Data Mining and Analytics, 4(3), 139–154. https://doi.org/10.26599/BDMA.2020.9020026

Ning, C., Zhang, S., and Li, L. (2009). Sustainable project management: A balance analysis model of effect. In Proceedings of the 2009 International Conference on Management and Service Science (pp. 1-4). IEEE. https://doi.org/10.1109/ICMSS.2009.5302334

Nisa, S. (2024). The environmental impacts of tourism on local communities in Pakistan.

Omran, A., and Kamran, H. W. (2018). The impact of climate change on our life: The question of sustainability. ISBN 978-981-10-7747-0.

O'Hagan, M., and Mangiron, C. (2013). Game Localization: Translating for the Global Digital Entertainment Industry. John Benjamins Publishing Company. https://doi.org/10.1075/btl.106.

Paskahlia Gunawan, A., and Tho, K. L. (2021). The Role of Local Content in Global Travel Platforms: An Analysis of Tourism Data. Journal of Travel Research

Paskahlia Gunawan, E., and Tho, C. (2021). Development of an application for tourism route recommendations with the dijkstra algorithm. Proceedings of 2021 International Conference on Information Management and Technology, ICIMTech 2021, 343–347. https://doi.org/10.1109/ICIMTech53080.2021.9534998

Raza, S., Khan, K., Qamar, B., Raza, S., Komal, A., Khan, B., and Syed, A. (2023). Understanding the influence of environmental triggers on tourists' pro-environmental behaviors in Pakistan's tourism industry. Journal of Tourism Futures. https://doi.org/10.

Roy, D., Dutta, M. (2022). A systematic review and research perspective on recommender systems. J Big Data 9, 59. https://doi.org/10.1186/s40537-022-00592-5.

Ricci, F., Rokach, L.,and Shapira, B. (2015). Recommender Systems Handbook. https://doi.org/10.1007/978-1-4899-7637-6

Sui, L., and Lin, L. (2020). Design of Household Cleaning Robot Based on Low-cost 2D LIDAR SLAM. 2020 International Symposium on Autonomous Systems, ISAS 2020, 223–227. https://doi.org/10.1109/ISAS49493.2020.9378863

Salman, S. (2024). Top Tourist Destinations in Pakistan. Kaggle. https://www.kaggle.com/datasets/naseeruddin444/top-tourist-destinations-in-pakistan

Association for Computing Machinery. (2007). RecSys '07: Proceedings of the 2007 ACM Conference on Recommender Systems: Minneapolis, MN, USA, October 19-20, 2007. Association for Computing Machinery.

Simuka, J., Marume, T., Mzembi, T., and Tagwi, S. (2024). Role of Technology to Improve the Effectiveness of University Canteen Systems. Journal of Business and Econometrics Studies, 1–6. https://doi.org/10.61440/JBES.2024.v1.13

Werthner, H., Stock, O., and Zancanaro, M. (2006). Futuring Travel Destination Recommendation Systems. In D. Fesenmaier, H. Werthner, & K. Wöber

(Eds.), Destination Recommendation Systems: Behavioral Foundations and Applications. CAB International. http://hdl.handle.net/20.500.12708/25382

Wang, J., Shan, T., and Englot, B. (2019). Virtual Maps for Autonomous Exploration with Pose SLAM. 4899–4906. https://doi.org/10.1109/IROS40897.2019.8967853

Wang, X., He, L., and Zhao, T. (2018). Mobile Robot for SLAM Research Based on Lidar and Binocular Vision Fusion. Chinese Journal of Sensors and Actuators, 31, 394–399. https://doi.org/10.3969/j.issn.1004-1699.2018.03.013

Xiang, Z., and Fesenmaier, D. R. (2021). Exploring the role of local experts in providing authentic travel information: A collaborative approach. Journal of Travel Research, 60(8), 1678-1692. https://doi.org/10.1177/00472875211004817.

Xiang, Z., and Fesenmaier, D. R. (2020). Business models and strategies of online travel agencies: Implications for online travel recommendation systems. Journal of Information Technology & Tourism, 22(1), 25-45. https://doi.org/10.1007/s40558-020-00164-4.

Xu, S., Fu, X., Cao, J., Liu, B., and Wang, Z. (2020). Survey on user location prediction based on geo-social networking data. World Wide Web, 23(3), 1621–1664. https://doi.org/10.1007/s11280-019-00777-8

Yin, H., Cui, B., Chen, L., Hu, Z., and Zhang, C. (2015). Modeling location-based user rating profiles for personalized recommendation. ACM Transactions on Knowledge Discovery from Data, 9(3). https://doi.org/10.1145/2663356

Zhou, W., Li, Y., and Chan, K. (2019). Digital Payment Methods and the Travel Industry: A Global Perspective. Tourism Management, 70, 378-391.

A. APPENDIX

A.1. Data Handling and Preprocessing

In this section, we walk through handling this dataset which is a crucial step towards any machine learning project. This is about handling missing values, renaming columns, encoding categories. The following code uses the input from a csv file of the name “Tourist Destinations.csv”. Normalize district names. It preserves consistency, as if we do not have any other naming than ‘Gilgit−Baltistan’, it’d replace it with ‘Gilgit-Baltistan’. It brings non numeric column (district) to numeric form. LabelEncoder is appllied to category section. It is convenient to use the data for developing machine learning models or to extract features. Transfromations are stored in label encoders for the mappings.

This process results in the actual outcome of this, the 'district' and the 'category' fields have become numerical within the data set, and said data is now ready for the subsequent analysis or for training the model.

```
df = pd.read_csv('/content/drive/MyDrive/Tourist Destinations.csv')


df['district'] = df['district'].replace({'Gilgit−Baltistan': 'Gilgit-Baltistan'})
label_encoders = {}
for column in ['category', 'district']:
le = LabelEncoder()
df[column] = le.fit_transform(df[column])
label_encoders[column] = le
```

A.2. Model Training

Machine learning model training shows how the technical method is used to develop the recommendation engine. The tourist destination is predicted from district, latitude, and longitude involving an AdaBoost classifier. To do this we need to first prepare our dataset by ‘encoding’ categorical variables such as district and category. The variables including the above mentioned have been used as features (X) and the target (y) is the tourist category. Then the data is split into training and testing sets of 70 percent for training and 30 percent for testing. Now, the model can use the feature that we provided to predict new data for category labels.

After that, it is trained on the model AdaBoost with 100 estimators. For classification tasks such as the one presented by this thesis, iteratively strengthening an already performing model and adapting it to misclassified instances of AdaBoost makes it special.

```
X = df[['category', 'district', 'latitude', 'longitude']]
y = df['category']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
model = AdaBoostClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

A.3. Budget Calculation and Suggestion

In this section, The generation of the model’s predictions, and calculated a budget from district selection. It gives the user details about how the system works. The code defines two important functionalities for the travel recommendation system, the budget calculation and recommendation . The calculate budget function assumes the budget has a known figure and multiplies the cost per night with this and adds both up to arrive at how much the user will spend on accommodation. After processing the user input, the recommendation function generates customized travel suggestions. This is based on the user's preference and have AdaBoost classifier help vote for the best tourist location, and have DataFrame for model prediction, followed by training of pre-fitted label encoders to encode the district and category into numeric distribution. The reliability of this mixed method allows users to receive an exact, cost budgeted estimate of the stay and recommendations personalized.

```
def calculate_budget(district, nights_stay):
per_night_cost = updated_district_budget_map.get(district, 2000

def predict(district, category, nights_stay):
district_encoded = label_encoders['district'].transform([district])[0]
category_encoded = label_encoders['category'].transform([category])[0]
 input_data = pd.DataFrame([{
        'category': category_encoded,
        'district': district_encoded,
        'latitude': 0,
        'longitude': 0
    }])

    predicted_point = model.predict(input_data)
    predicted_point_category = predicted_point[0]

    budget = calculate_budget(district, nights_stay)
    return f"Estimated Budget: {budget} PKR"
```

A.4. Predictions on Test Set

This section generates predictions using the AdaBoost classifier model trained on the dataset. The model predicts values for X_test to compare with actual labels (y_test), which serves as a foundation for evaluating model performance.

```
y_pred = model.predict(X_test)
```

A.5. Confusion Matrix

The accuracy of the classifier is shown by confusion matrix which breaks down correct and incorrect predictions for every category. Using seaborn, the matrix is plotted as a heatmap, expressing areas of high and low prediction accuracy.

```
cm = confusion_matrix(y_test, y_pred)

plt.figure(figsize=(10,7))

sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')

plt.xlabel('Predicted')

plt.ylabel('Actual')

plt.title('Confusion Matrix')

plt.show()
```

A.6. Classification Report

This report includes metrics like precision, recall, F1-score, and support for each class. It is a comprehensive evaluation of the model's performance, highlighting strengths and areas for improvement in predictions.

```
report = classification_report(y_test, y_pred)

print("Classification Report:\n", report)
```

These sections discuss the evaluation results and the model performance in predicting each category. The code give information on how the model is performing with respect to the classification report metrics that help determine predictive ability of the model.