Table of Contents

1. Introduction

Recommender Systems (RSs) are the software that learns about the interests of the users and provides recommendations that the user would find most useful [1]. The recommendation system is helpful for those people who are not good at comparing similar items provided by different websites. Most of the RSs are based on user ratings of the items [1]. Recommendation System is a current favourite topic among the researcher as there is ample information is available on the web for hotels, movies, books, jobs, music, video, and many more [1][10]. So, the recommendation system is very useful nowadays as it allows users to access the information in which they are interested. If we are planning for something, then it's human tendency that we first search on the web related to that and try to find out best option from the different available options for that item, for example, if we are planning a trip first, we will search for the hotel in that particular area and read review for some hotels and most likely based on the reviews we decide a hotel for a stay. So, to fulfill this need recommendation system comes in a picture.

Recommendation System provides help to users to find out proper information about an item at the right time [2]. Recommendation systems are developed mostly based on Collaborative filtering, Content-based filtering, Hybrid filtering and Knowledge-based filtering [2]. Below we have given a brief introduction on Collaborative and Content-based filtering methods:

- **Collaborative Filtering:** RSs based on this technique, suggests items based on how other similar users like that item. In a simple word, a collaborative filtering method filters the information about the user's interest by gathering information from other similar users [8].

- **Content-based filtering:** This method, also known as cognitive filtering. RSs based on the content-based filtering use the data provided by the user such as ratings or reviews of items; based on the review of an item user profile is generated, and then the item is recommended by RSs using user profile and item description [2][9][7].

So, overall, Recommender Systems uses different filtering methods to recommend the best items to the users. Currently, many recommender systems are based on content-based filtering. [3] Nowadays, with the advancement of technology, there are ample e-commerce websites that provide different options for similar items like electronic items at Amazon, hotel options at Tripadvisor and many more. So, if we are selecting something from e-commerce websites, we generally first go for the reviews of that particular item that we want. So many recommendation systems use reviews given by users to suggest an item for

users. Online reviews have become an essential part of the user decision making process about an item, so for that reason, almost all online sites provide an option to share your experience about an item on their website for example, E-Commerce and Hotel Websites gives you an option to write a review or share your experience for hotel that you have visited [4]. Data provided by online reviews is rich in information about user's preferences, user likes and dislikes as well as provides the information about whether that item is good or not [4]. So, using the information provided by user reviews, we can develop better recommendation systems to attract more customers and suggest to them a better recommendation of products and this will ultimately helpful for sellers as well to increase their profit.

1. 1 Problem Statement:

Currently, many existing recommendation systems available in the market, but in that, we might be experiencing scalability and inefficiency problems while discovering and working with large-scale data. They have not considered the customer's various preferences and cannot meet the user's personalized requirements. In that system user is not the representative of the overall statistics. Even in the recent rating approaches allow users to specify their reviews but does not use this rating to discover recommendation. So, for the better recommendation purpose, a review-based hotel recommendation system can work on those user reviews and suggest hotels based on a variety of input criteria. Online reviews are generally available in both forms, such as rating and text-based comments about product/item/hotel. But, Most of the systems ask users to give the rating in low to large scale for the hotels and how they like their services and then compute overall rating for a hotel. Our system will mostly work based on the content available in on textual user reviews because it can be helpful when the user does not provide enough ratings or false ratings.

Existing System:

The existing system proposed by author used a collaborative filtering technique to give a recommendation to the user. In this existing hotel recommendation system, lots of information regarding client behaviour is given. These incorporate what the user is looking for and predict which kind of hotel a user is going to book. This system has its in-house algorithms to form a hotel cluster and this cluster works as good identifiers and it can suggest new hotels but it doesn't have past records of hotels and user reviews.

Drawbacks of the existing system:

- In this existing system, it doesn't store historical data of user reviews so, the system cannot suggest suitable hotels to the user.

- The system is using collaborative filtering, and it requires some minimum number of users to rate a new thing before that thing can be recommended.
- The existing system needs structured data, and it takes time.

1. 2 Project Context:

This project will explore different machine learning methods through a literature review on hotel recommendation system or general recommendation systems and determine which machine learning method is good at providing a better recommendation for hotels, and we will try to build a better recommendation system using the content of the user's reviews and ratings provided by users.

1. 3 Motivation:

- The recommender system's main aim is to proposed items such as hotels, and that is the potential to be liked by the users [11].
- It is the best way to promote everything to the right users because now online reviews are becoming the real evaluation for anything [11].
- Customers are willing to share their opinion on social media, and this can provide an opportunity for the hotel or any business staff to improve their services.
- The recommendation system can help sellers to learn their user's preferences [11].

2. Project Overview:

This section defines the objectives, project scope, and target deliverables.

2. 1 Objectives:

- The main aim of this project is to develop a better hotel recommendation system that can suggest more accurate hotels based on the reviews provided by users.
- Implement new machine learning methods to extract information from user's reviews on different hotels such as Natural Language Processing for text mining; sentimental analysis can be used for this.
- Improving user experience by providing better recommendations by the recommender system.

2. 2 Deliverables:

- Analyzing literature review on existing recommendation system methods so that these findings can be used to develop a better hotel recommender system.
- Implementation of machine learning techniques/algorithms for recommendations.
- Project Report.

2. 3 Project Scope:

- In this project, the main work is involved around scrapping data from different hotel websites, pre-processing of data and developing a hotel recommendation system until it completed.
- Through the literature review, we will examine existing methods for hotel recommendations with different machine learning algorithms/methods.
- Design a recommender system that can recommend hotels based on the content of user's reviews

2. 4 Project Significance:

- This project is focusing on developing a hotel recommender system that uses reviews provided by customers for different hotels, and by performing text analysis on the collected data and using machine learning algorithms, we will try to build a better hotel recommender system that can improve user experience.
- If this developed system is implemented on a large scale and provides better performance than existing systems, then many users can get benefits from it and get a better user experience for hotel recommendations.

3. Solution Overview:

Our main goal is to recommend the most suitable hotels to a user based on the reviews provided for hotels. In general, our first step is to identify important information form the user's reviews because text-based comments contain many sentences. To build a recommender system first, we need data to play with it; for that, we will first do web scrapping of user reviews from different hotel websites, which will be used later. Online reviews mostly consist of ratings of item/hotel, and textual comments [6], ratings provided by users are easy to process but text-based reviews are mostly in a natural language without any formation [3][6]. Customer writes their reviews on their own words based on what they

experienced. While recommender systems need structured well-formed data to build a model and to provide proper recommendations, but textual reviews are very useful in order to build a better recommendation system. So, processing will be required to convert user's reviews to meaningful information i.e. unstructured data to useful data.

Text mining is used to convert unstructured data into meaningful data [6]. From the text-based reviews of users, we need to extract useful information, which involves sentiment analysis of text and opinion mining. Sentiment analysis is one of the parts of Natural Language Processing (NLP). To fetch or extract emotions from the text, this analysis is being used. Probably after gathering sufficient data for our project, we will do sentiment analysis of the collected data using python, for that will use "NLTK" module of python, Scikit-learn library for this NLP method and predict whether customer is satisfied or not by that hotel based on review given by customer, which will be useful for extracting features and for recommendation modelling.

We will develop a specialized recommender system for hotels and which will not only focus on various rating parameters but also considering text-based comments on the client's reviews.

4. Dataset:

We are planning to collect user reviews for hotels through web-scrapping from different websites. [12] Web-Scrapping is a method to extract data from websites, you can extract a large amount of data using this technique, and it allows you to store the data in your local machine and you can then use that data for further analysis. Different software or tools are available that can perform web-scraping or we can run some python code through "Jupyter Notebook" and collect the data in a local file or store them in the database in tabular format. Currently, we will most likely collect the data by using the python code, pandas and will store it in the local machine. We are not sure about the size of datasets yet but will try to collect as much as user's reviews from different websites and make sure that we have sufficient data for the hotel recommendation system. We will store our data as in .CSV format currently.

To collect the user's review for hotels, we will use below-mentioned websites:

- https://www.tripadvisor.ca/
- https://www.booking.com/
- Trivago etc.