A PROJECT ON

“HOTEL RECOMMENDATION SYSTEM”

SUBMITTED IN
PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR THE COURSE OF
DIPLOMA IN BIG DATA ANALYSIS

SUNBEAM INSTITUTE OF INFORMATION TECHNOLOGY, PUNE

Submitted By:

Deepak Chandra (83830)

Aniket Jadhav (83932)

Mr.Nitin Kudale
CEO

Mrs.Manisha Hingne
Course Coordinator

CERTIFICATE

This is to certify that the project work under the title ‘House Price Prediction in Australia’ is done by Deepak Chandra & Aniket Jadhav in partial fulfillment of the requirement for award of Diploma in Big Data Analysis Course.

Mr. Aniket P
Project Guide

Mrs. Manisha Hingne
Course Coordinator

Date: 13/08/2024

ACKNOWLEDGEMENT

A project usually falls short of its expectation unless aided and guided by the right persons at the right time. We avail this opportunity to express our deep sense of gratitude towards Mr. Nitin Kudale (CEO , Sunbeam, Pune) and Mrs. Manisha Hingne (Course Coordinator, Sunbeam ,Pune) and Project Guide Mr. Aniket P.

We are deeply indebted and grateful to them for their guidance, encouragement and deep concern for our project. Without their critical evaluation and suggestions at every stage of the project, this project could never have reached its present form.

Last but not the least we thank the entire faculty and the staff members of Sunbeam Institute of Information Technology, Pune for their support.

Deepak Chandra

DBDA March 2024
Batch, Sunbeam Pune

Aniket Jadhav

DBDA March
2024 Batch,
Sunbeam Pune

TABLE OF CONTENTS

1. Introduction

1.1 Introduction And Objectives:

The hospitality industry has grown exponentially over the years, with numerous hotels vying for the attention of travelers. In this competitive landscape, offering personalized recommendations to users has become a crucial factor in enhancing customer satisfaction and driving bookings. The goal of our hotel recommendation system is to assist users in finding the perfect accommodation by analyzing user preferences, past reviews, and other relevant data.

1.2 Why this problem needs To be Solved?

In today's digital age, the abundance of choices in the hospitality industry can be overwhelming for travelers. With thousands of hotels to choose from, selecting the right accommodation that meets an individual's specific needs and preferences can be a daunting task. A robust hotel recommendation system is crucial for simplifying this decision-making process, ensuring that users are presented with personalized and relevant options. In our project, we aim to develop a recommendation system that leverages user reviews, ratings, and hotel features to suggest the most suitable hotels for each user. By understanding user preferences and analyzing past behavior, the system can provide tailored recommendations that enhance the user experience.

1.3 Dataset Information

page_in_pages.csv:

The dataset contains 17 columns:

1. **Hotel_Address:** Address of the hotel.
2. **Additional_Number_of_Scoring:** Additional number of scoring metrics used in reviews.
3. **Review_Date:** Date when the review was posted.
4. **Average_Score:** Average score of the hotel based on reviews.
5. **Hotel_Name:** Name of the hotel.
6. **Reviewer_Nationality:** Nationality of the reviewer.
7. **Negative_Review:** Text of the negative review provided by the user.
8. **Review_Total_Negative_Word_Counts:** Total number of words in the negative review.
9. **Total_Number_of_Reviews:** Total number of reviews the hotel has received.
10. **Positive_Review:** Text of the positive review provided by the user.
11. **Review_Total_Positive_Word_Counts:** Total number of words in the positive review.

12. **Total_Number_of_Reviews_Reviewer_Has_Given:** Total number of reviews the reviewer has given.
13. **Reviewer_Score:** Score given by the reviewer to the hotel.
14. **Tags:** Tags associated with the review, indicating the type of trip, type of room, etc.
15. **days_since_review:** Number of days since the review was posted.
16. **lat:** Latitude of the hotel.
17. **lng:** Longitude of the hotel.

2. Problem Definition and Algorithm:

2.1 Problem Definition

The challenge we are addressing involves developing a system that can accurately recommend hotels to users based on their preferences and past behavior. We have gathered a dataset containing user reviews, ratings, and various features of hotels from different locations around the world. Our task is to utilize this data to build a recommendation system that can suggest the most suitable hotels to users.

In this project, our metrics of interest include Mean Square Errors(MSE), which will help us evaluate the effectiveness of our recommendations.

2.2 Algorithm Definition

User-Item Interaction Matrix: The user-item interaction matrix is a fundamental concept in recommendation systems. It represents the interactions between users and items (in this case, hotels) in a matrix format, where rows correspond to users, and columns correspond to items. The entries in the matrix typically indicate the rating or interaction between a user and a hotel, such as a review score or a binary indicator of whether a hotel was booked. This matrix serves as the foundation for many collaborative filtering techniques, where the goal is to predict missing entries (e.g., which hotels a user might like but hasn’t yet interacted with) by identifying patterns in the data.

Item-Based Collaborative Filtering: Item-based collaborative filtering is a method that recommends items to users based on the similarity between items. In the context of a hotel recommendation system, this approach

identifies hotels that are similar to those a user has previously rated or interacted with and suggests these similar hotels to the user. The similarity between hotels is typically computed using metrics like cosine similarity or Pearson correlation, based on user ratings or interaction data. Item-based collaborative filtering is particularly effective when users tend to interact with items that have well-defined patterns of similarity, such as hotels with similar amenities or in similar locations.

User-Based Collaborative Filtering: User-based collaborative filtering recommends hotels to a user by finding other users with similar preferences and suggesting hotels that those similar users have liked. The algorithm calculates the similarity between users using techniques such as cosine similarity or Pearson correlation, based on their interaction histories. Once similar users (neighbors) are identified, the system predicts the hotels that the current user might like by aggregating the preferences of these neighbors. This method is effective when users with similar tastes tend to rate or interact with similar hotels, making it a powerful approach for personalized recommendations.

Content-Based Filtering: Content-based filtering recommends hotels by analyzing the features of the hotels themselves, rather than relying on user interactions. This approach suggests hotels that are similar to those a user has liked in the past, based on characteristics such as location, amenities, price range, and customer reviews. The algorithm creates a user profile based on the attributes of the hotels the user has interacted with and then matches this profile against the attributes of other hotels in the dataset to generate recommendations. Content-based filtering is particularly useful when there is rich information available about the items (hotels), and it allows for recommendations even when user interaction data is sparse.

3. Experimental Evaluation:

3.1 Methodology:

The objective of this project is to recommend hotel according to user input. The data is taken from booking.com dataset and put into a dataframe so that preprocessing and analyzing can be done.

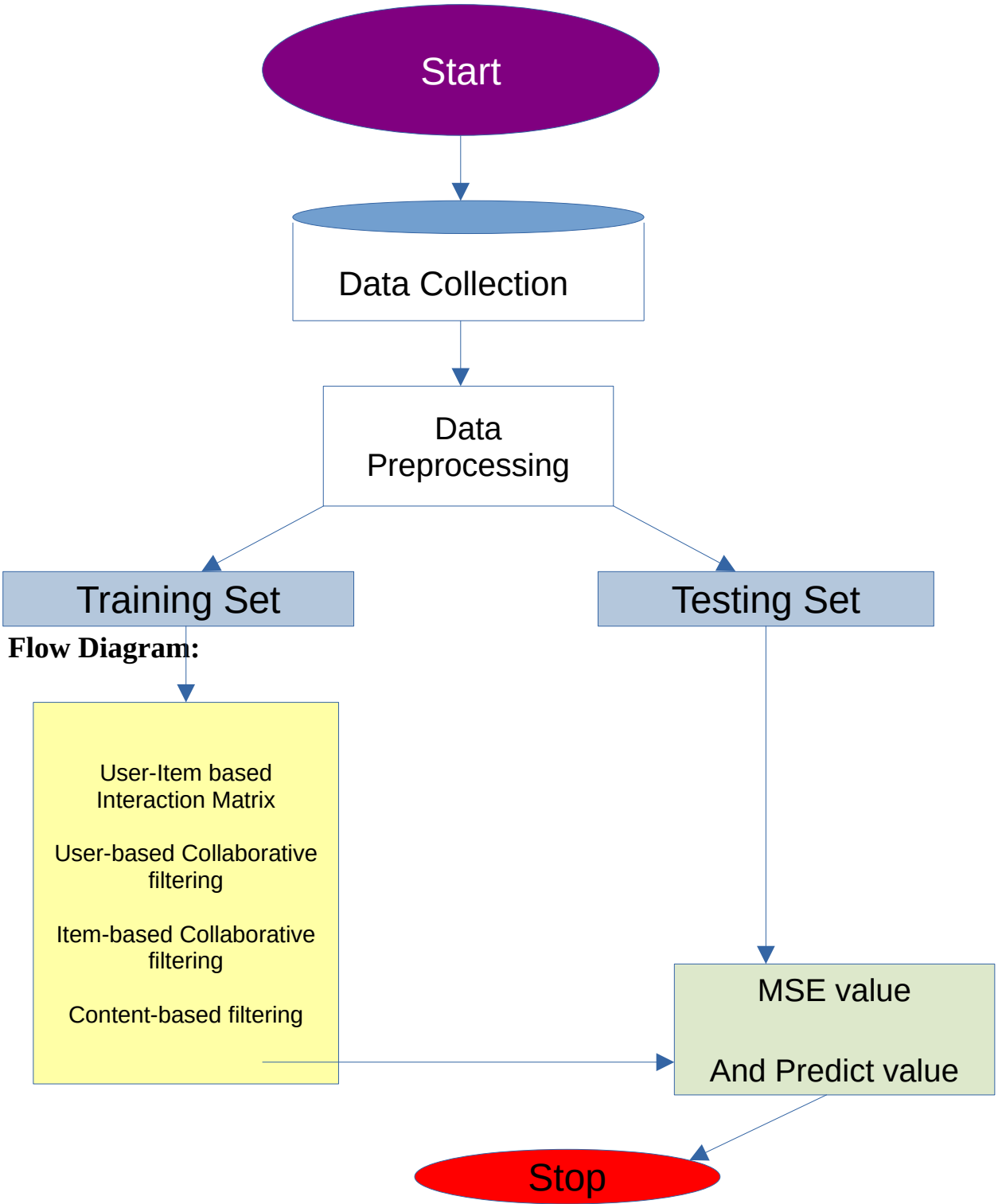
Loading in raw Data:

```
df = pd.read_csv('booking_review_new2.csv')
df.head()
```

Feature Engineering and Processing :

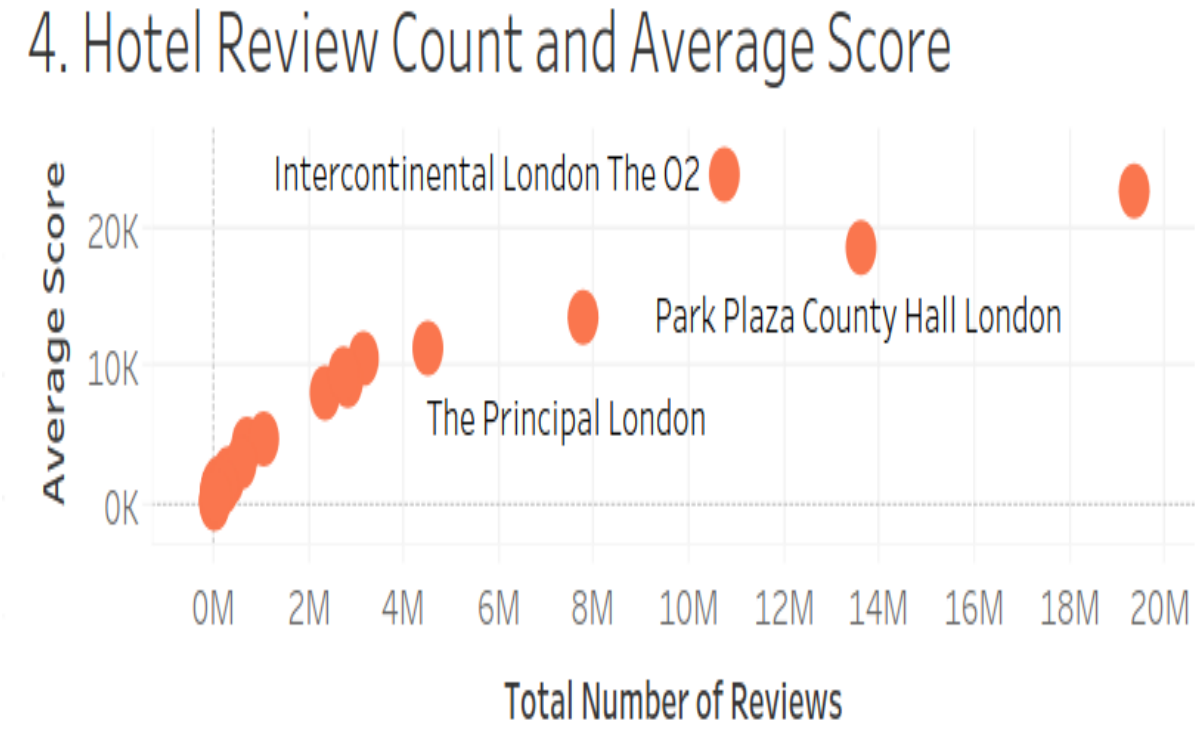
```
# Combine features to create a unified feature for text processing
  train_data['combined_features'] = train_data['Hotel_Name'] + ' ' +
train_data['Positive_Review'] + ' ' + train_data['Negative_Review']


# Convert text data into numerical data using TF-IDF Vectorizer
  tfidf = TfidfVectorizer()
  tfidf_matrix_train = tfidf.fit_transform(train_data['combined_features'])

# Apply Truncated SVD to reduce dimensionality
  n_components = 100  # Number of components to keep
  svd = TruncatedSVD(n_components=n_components, random_state=42)
  tfidf_matrix_reduced = svd.fit_transform(tfidf_matrix_train)

# Compute cosine similarity on the reduced matrix
   content_similarity_train = cosine_similarity(tfidf_matrix_reduced)
   content_similarity_df_train = pd.DataFrame(content_similarity_train,
   index=train_data['Hotel_Name'], columns=train_data['Hotel_Name'])
```

Flow Diagram:



Flow Diagram:

3.2 Exploratory Data Analysis

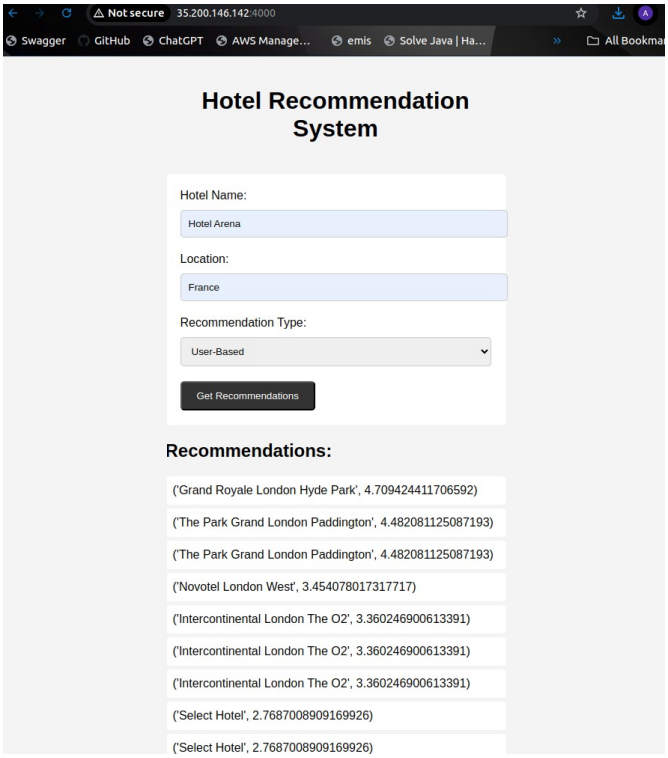


This graph shows average score according to total number of reviews and the name of the hotels are mentioned on the graph

4. GUI:

GUI is made using Flask framework. **Flask** is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools

Link : http://35.200.146.142:4000/



5.GitHubLink:

https://github.com/AniketJ5/Hotel_Recommendation_System.git

6. Future work And Conclusion

6.1

Future Work:

- **Incorporation of Advanced Machine Learning Techniques:**

    - Explore the use of deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to analyze user reviews, images, and sequential booking behavior to capture more complex patterns in user preferences and hotel features.
- **Integration of Location-Based and Contextual Data:**
    - Incorporate location-based data such as proximity to tourist attractions, public transport, and local events to enhance the relevance of recommendations. This could also include contextual data such as the season or time of year to tailor recommendations based on travel trends.
- **Utilization of Sentiment Analysis and Social Media Data:**
    - Leverage sentiment analysis on user reviews and social media data to better understand customer satisfaction and trends. Analyzing user sentiment can help in fine-tuning recommendations and identifying hotels that meet specific user expectations.
- **Implementation of Real-Time Recommendation Systems:**
    - Develop real-time recommendation capabilities that update suggestions based on live data, such as current hotel availability, price fluctuations, or last-minute booking trends. This would involve the use of streaming data processing frameworks to handle and analyze real-time data efficiently.
- **Personalized Recommendation Strategies:**
    - Enhance personalization by using advanced customer segmentation and profiling techniques. This could involve analyzing user behavior across multiple platforms or integrating data from loyalty programs to deliver highly personalized hotel recommendations.

6.2 Conclusion:

Features like Hotel Name, Positive Review, Negative Review, and Reviewer Score are likely to be significant predictors of user preferences and hotel recommendations.

- If Reviewer Score has a high feature importance score, it suggests that hotels with higher overall ratings are more likely to be recommended to users, aligning with common trends where users tend to prefer highly rated hotels.
- Hotel location and user nationality are major contributing factors for predicting user preferences, as they reflect the importance of geographical and cultural relevance in hotel choice.
- By analyzing the Total Number of Reviews, you might conclude whether a hotel's popularity is increasing or decreasing, revealing trends in customer satisfaction and brand loyalty.
- Sentiment analysis on Positive and Negative Reviews can reflect the overall satisfaction level and common pain points of users, indirectly influencing hotel recommendations by highlighting what aspects of the hotel experience are most valued or criticized.

- Among the trained models for recommending hotels, item-based collaborative filtering algorithm performs the best, providing accurate and personalized recommendations by leveraging the similarities between hotels.