CSE4077- Recommender Systems

J Component – Project Report

Airbnb recommendation system

By

19MIA1061	**ARAVINDAN T R**
19MIA1047	**MOHAMMED IMRAN Z**
19MIA1082	**ALAGARSAMY N**
19MIA1062	**ABINANDHAN KUMAR T S S**

M.Tech CSE with Specialization in Business Analytics

Submitted to

Dr.A.Bhuvaneswari,
Assistant Professor Senior,
SCOPE, VIT, Chennai

School of Computer Science and Engineering

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

August 2022

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computing Science and Engineering

VIT Chennai

Vandalur - Kelambakkam Road, Chennai - 600 127

FALL SEM 22-23

Worklet details

Programme	Computer Science and engineering with specialization in business analytics	
Course Name / Code	Recommender Systems	
Slot	E1+TE1	
Faculty Name	BHUVANESWARI	
Component	J – Component	
J Component Title	Airbnb recommendation system	
Team Members Name | Reg. No	ARAVINDAN T R	19MIA1061
	MOHAMMED IMRAN Z	19MIA1047
	ALAGARSAMY N	19MIA1082
	ABINANDHAN KUMAR T S S	19MIA1062

ABSTRACT

Airbnb is an online marketplace and hospitality service where people can rent short-term lodging such as apartments, hostel beds, hotel rooms and cottages. People can also organize or participate in holiday activities and experiences such as walking tours, concerts, workshops and restaurant dining. There are more than 4 million accommodation listings on Airbnb in 191 countries and 65000 cities, with over 260 million check-ins facilitated

Airbnb can be accessed via its website or mobile apps. Accommodation listings are generated when users search by destination and use filters such as Dates, No. of Guests, Home Type, Price and Trip Type. Airbnb is popular among travellers, especially those who are budget-conscious, because of its various advantages. For example, travellers have the option to stay in an entire apartment which offers greater flexibility compared to a hotel room. They are also able to have a more authentic travel experience by staying in a local's home, and prices are generally lower than hotels .

One downside of relying on Airbnb for travel planning, however, is that listings can be fully booked very quickly, especially those in desirable locations and during peak travel periods. In addition, travellers would have to look through reviews of listings carefully to ensure safety and security as well as to be better informed about the amenities provided by the host.

This project aims to build an recommendation system based on the polarity scores of the user review using sentiment analysis. It makes use of datasets provided by Inside Airbnb, which are sourced from publicly available information from the Airbnb site. These datasets include detailed information on listings and reviews by Airbnb users, for a number of cities and countries. The project focuses on accommodation listings in London.

1. Literature Survey

Sl no	Title	Author / Journal name / Year	Technique	Result
1	Airbnb Price Prediction Using Machine Learning and Sentiment Analysis	Pouya Rezazadeh Kalehbasti, Liubov Nikolenko, Hoormazd Rezaei	K-means Clustering, Support Vector Regression, Neural Network, Gradient Boosting.	Support Vector Regression (SVR) performed the best and produced an R2 score of 69% and a MSE of 0.147 (defined on ln(price)) on the test set.

2	Airbnb Price Prediction using sentimental analysis	Peilu Liu	Linear Regression, Grdient boost, Neural network, SVR.	Among the models tested, Support Vector Regression (SVR) performed the best and produced an R2 score of 69% and a MSE of 0.147 (defined on ln(price)) on the test set.
3	Applying Deep Learning To Airbnb Search.	Malay Haldar & Moose Abdool	CNN, LTSM, DNN	Overall, we found this led to an increase in the diversity of our search results, along with a +0.4% global booking gain in an online A/B test.
4	Realtime personalization using embeddings for search ranking at airbnb	Kamelia aryafar, Devin Guillory, and Liangjie Hong	Embedded models	From the results of SLR, we analyze that various MBSE activities are simultaneously researched to provide a complete development solution for embedded systems.
5	Self-Supervised learning on graph	Kadhar Moidheen	GCN	The model which they used got the accuracy of 81.32%, which is cora, then 71.43% which is citeseer, then they got 71.28% for pumped.

6	A Hotel Recommendation System Based on Reviews, 2016 Fourth International Symposium on Computing and Networking (CANDAR)	Koji Takuma, Junya Yamamoto, Sayaka Kamei, Satoshi Fujita	method to extract the preference of review contributors from a collection of reviews.	result of questionnaire-based evaluations indicates that our proposed method can recommend hotels that matches the user preference.
7	Integrating contextual sentiment analysis in collaborative recommender systems	PLOS ONE 2021, Nurul Aida Osman, Shahrul Azman Mohd Noah, Mohammad Darwich, Masnizah Mohd	a sentiment-based model with contextual information for recommender system was proposed.	Results showed that the proposed contextual information sentiment-based model illustrates better performance as compared to the traditional collaborative filtering approach.
8	Reviews Sentiment analysis for collaborative recommender system	Alia Karim Ahmed Bahaa, 1st International Conference on Engineering and Computing, 2017 (ICEC2017) 2017	Sentiment analysis system implemented using NLP techniques with machine learning to predict user rating form his review	Sentiment analysis success in predicting user satisfaction or dissatisfaction by classifying reviews into either positive or negative. This approach could compensate the deficiency in user rating about an item in recommender system (data sparsity).

9	Collaborative Filtering Recommender System: Overview and Challenges	Hael Al-bashiri, Mansoor Abdullateef Abdulgabber, Awanis Romli, Fadhl Hujainah, Journal of Computational and Theoretical Nanoscience 2017	defined the main challenges which have clearly impact on the performance and accuracy of CF recommender system.	This paper summarizes the limitations of the existing methods and recommendations.

2. Dataset and Tool to be used (Details)

We will be using the **Boston Airbnb open data** dataset from Kaggle

Listings.csv: It contains full descriptions and average review score

Reviews.csv: It contains unique id for each reviewer and detailed comments

Calendar.csv: It contains listing id and the price and availability for that day

Tools: Python, Google Colab

3. Algorithms / Techniques description

It consists of 3 phases:

(a) Pre-processing

- Do the necessary text pre-processing steps for each reviewer's comments (including language detection)
- Using NLTK's Vader lexicon, calculate the compound polarity of each comment

(b) Build a recommendation engine – Collaborative filtering and Sentiment Analysis

- Once we have the polarities for each reviewer/listing pair, build a utility matrixwith reviewer_id on one axis and listing_id on the other
- Predict all polarities (fill in the Nan values of that matrix) using SVD

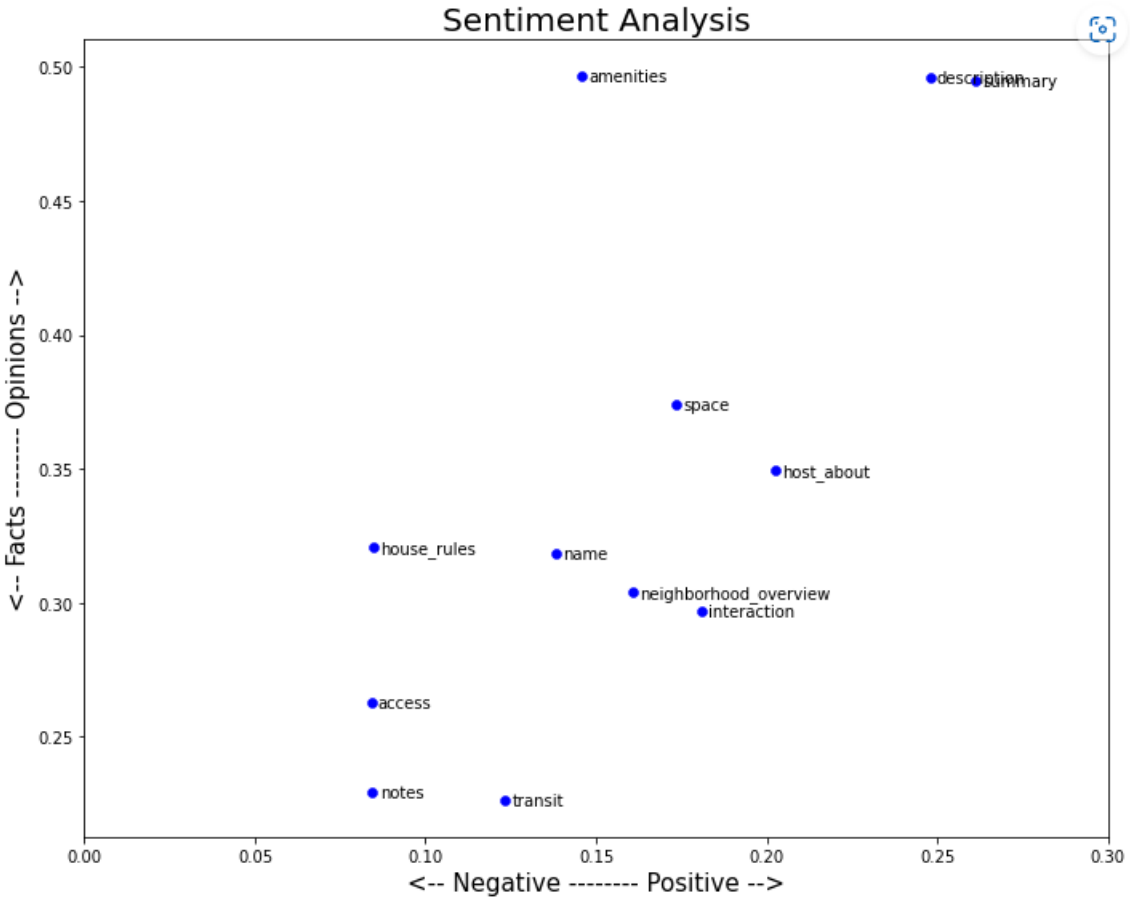
(c) Generate Recommendations

i. We can chose the number of top recommendations we want to give a particularuser, sorted according to their predicted polarity

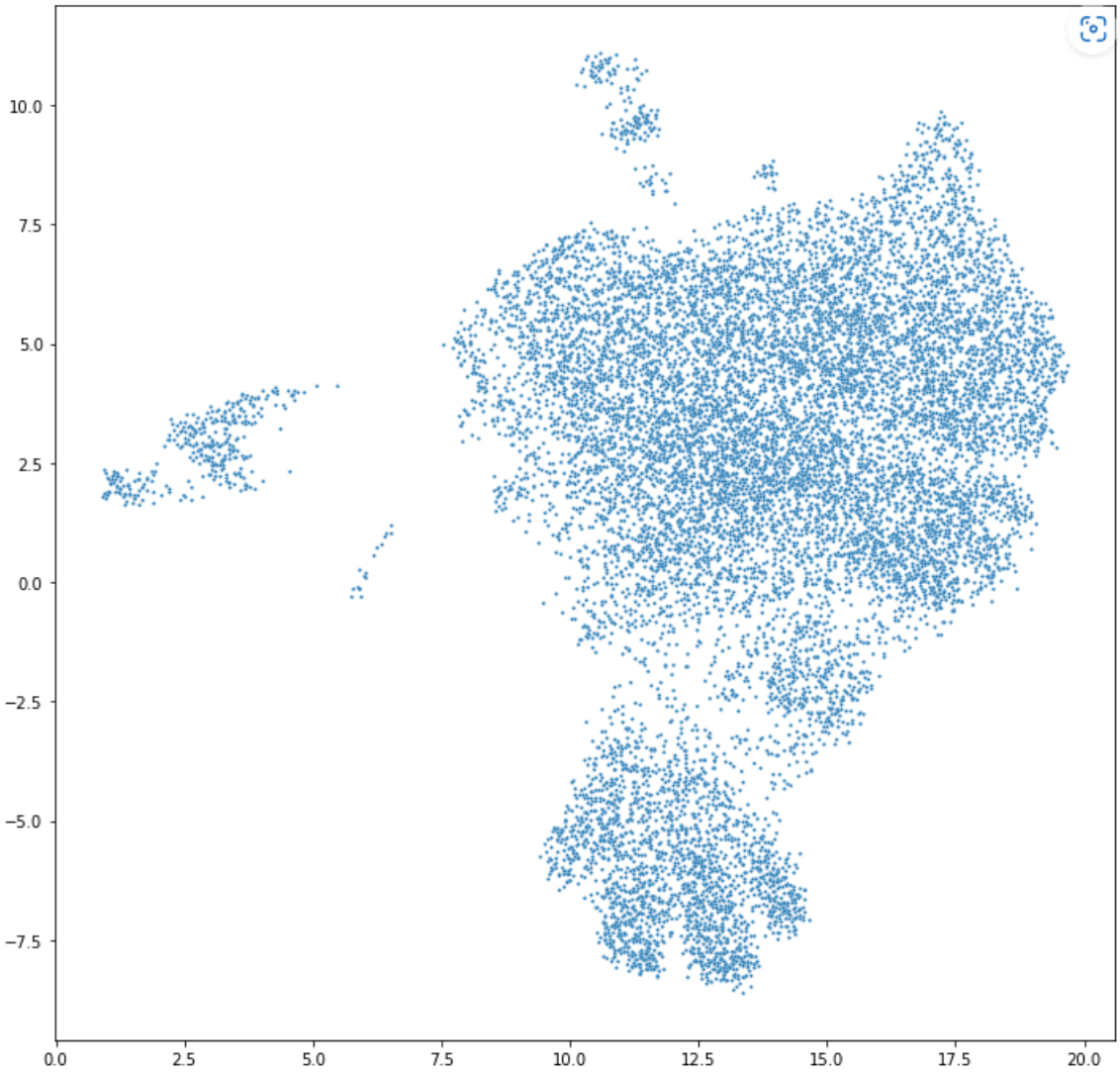
4. Implementation Details

- Performed Data cleaning like missing null values, encoding categorical and ordinal features.
- Implemented Tokenization and lemmatization for sentiment analysis.
- Calculated polarity and subjectivity for user reviews
- Converting all text features to numerical features
- Visualized the embeddings of the data
- Found cluster labels for each review

5. Results and Discussion



- Airbnb listings tend to be positive when it comes to descriptions and summaries.
- This makes sense, hosts want to encourage people to stay at their Airbnb and having a positive description is beneficial.
- However, these descriptions tend to be grounded in opinion
- Interesting to note that amenities are considered very opinionated.
- One would expect that amenities would be more grounded in facts.



- From the embedding plot, we can see that the dataset has different clusters
- We performed clustering to find the cluster labels for each and every review
- We will use the cluster labels and the calculated features to recommend similar Airbnb listing using Cosine Similarity

	id	cluster_label	latitude	longitude	neighbourhood_cleansed	zipcode	property_type	room_type	accommodates	bathrooms	bedrooms	beds	bed
0	33159143	2	32.91736	-117.07635	Scripps Ranch	92131	House	Private room	1	1.0	1.0	1	Re
1	17138468	3	32.84067	-117.27443	La Jolla	92037	Apartment	Entire home/apt	1	2.0	2.0	3	Re
2	21898446	3	32.79797	-117.24250	Pacific Beach	92109	Townhouse	Private room	1	1.0	1.0	1	Re
3	25948680	3	32.77545	-117.05923	College Area	92120	Apartment	Entire home/apt	1	1.0	1.0	1	Re
4	1756516	2	32.84619	-117.27558	La Jolla	92037	Condominium	Private room	1	1.0	1.0	1	Re

6. References

[1] Lovedeep Singh (2020). Fake News Detection: a comparison between available Deep Learning techniques in vector space, 4th Conference on Information & Communication Technology (CICT), Chennai, India, 2020, pp. 1-4, doi: 10.1109/CICT51604.2020.9312099.

[2] Al-Ghuribi S. M. & Mohd Noah S. A., "Multi-criteria review-based recommender system—the state of the art", in IEEE Access, 7, pp. 169446–169468, 2019.

[3] Osman N. A. & Mohd Noah S. A., "Sentiment-based model for recommender systems" in Proceedings of the Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), 2018.

[4] Ghabayen A. S., Mohd Noah S. A., "Using tags for measuring the semantic similarity of users to enhance collaborative filtering recommender systems," in International Journal on Advanced Science, Engineering and Information Technology, vol., no. 6, 2063–2070, 2017.

[5] Darwich M., Mohd Noah S. A. and Omar N., "Automatically generating a sentiment lexicon for the Malay language," in Asia-Pacific Journal of Information Technology and Multimedia, vol. 5., no. 1, 2016.

[6] Darwich M., Noah S. A. M., and Omar N., "Minimally-Supervised Sentiment Lexicon Induction Model: A Case Study of Malay Sentiment Analysis," Multi-disciplinary Trends in Artificial Intelligence—11th International Workshop, MIWAI 2017, Proceedings. Springer Verlag, Vol. 10607 LNAI, pp. 225–237, 2017.