AIRLINE RECOMMENDATION SYSTEM

By

Bharat Chandra Mokkapati

Harshit Shrimali

Krishna Sai Biradar

Pardha Saradhi Mogilipaka

Under Supervision of

Dr. Tony Diana

For

Spring-2022

ABSTRACT

To alleviate the problem of information overload, which has become a possible problem for many internet users, it is necessary to filter, prioritize, and efficiently communicate critical information on the Internet, where the number of alternatives is overwhelming. Recommender systems solve this problem by looking through a massive amount of dynamically produced data to provide personalized content and services. For example, when a passenger wants to travel, no site takes user preferences and recommends the flight based on his criteria. Therefore, a friendly personalized flight recommendation system is necessary, which can give importance to user preference and recommend the flight to choose from the options and travel according to his needs. We have used different techniques like collaborative filtering, content-based filtering, and hybrid filtering to generate the appropriate results.

INTRODUCTION

Nowadays, people are presented with many options, which often leads to confusion about which product would best meet their needs. As a result, the necessity for a system that could facilitate these selection criteria while also removing the masses, i.e., eliminating the problem of choosing a single choice among numerous choices issue was recognized, and the ultimately recommender systems of today's modern world were established. In today's modern world, the mode of traveling got completely changed. Nowadays, People want to travel fast and quickly. Travelers decide based on their preferences factors (airlines, classes, luggage fees, plane types, and ticket regulations) rather than merely on price or time. As a result, suggesting flights to customers has become an important consideration. Therefore, the recommendations are typically based on the consumers' specific travel preferences, which may be learned from their previous ratings or buying behavior. In this project, we have analyzed historical data of user ratings over five years. In our systems, the user selects the origin country and determines top priorities like inflight services, food, beverages, cabin staff ratings, etc. Then, the system analyses and evaluates the inputs and recommends the airline to the customer.

LITERATURE REVIEW

1. Recommendations on Airline Dataset System using Map Reduce and Mahout

Consumers find it challenging to choose from the numerous distinct alternatives offered in the market as diversity grows throughout every business. The company must present the most refined advised item based on the customer needs, product ratings, and client preferences. We examine a significant quantity of accessible data to make these recommendations as feasible as close to the consumers' choices. Recently, various approaches for building recommendation systems have been developed, utilizing either collaborative filtering, content-based filtering, or hybrid filtering. Group Lens is a news-based architecture that employs collaborative methods in assisting users in locating articles from a massive news database. Ringo is an online social information filtering system that uses collaborative filtering to build users' profiles based on their ratings on music albums. Amazon uses topic diversification algorithms to improve its recommendation. Despite the success of these two filtering techniques, several limitations have been identified. Some of the problems associated with content-based filtering techniques are limited content analysis, overspecialization, and data sparsity. Also, collaborative approaches exhibit cold-start, sparsity, and scalability problems. These problems usually reduce the quality of recommendations. To mitigate some of the identified issues, hybrid filtering, which combines two or more filtering techniques in different ways to increase the accuracy and performance of recommender systems, has been selected as a methodology. These techniques combine two or more filtering approaches to harness their strengths while leveling out their corresponding weaknesses.

2. A Comparative study on Airline Recommendation System Using Sentimental Analysis on Customer Tweets

Numerous areas require a recommendation, and it has become a prominent research topic. Understanding users' unique past behavior by assessing the tweets posted by them is significant to understanding recommender systems. The tweets are evaluated as 'positive,' 'negative,' or 'neutral' based on terms and overall context. Afterward, the score for each airline is calculated to rank each airline with respect to one another. Then, we applied the Naïve Bayes algorithm. This allowed us to count the number of positive, negative, and neutral tweets related to airlines. Individual scores for each airline were determined using both Naïve Bayes and sentimental analysis, and a comparison study is conducted to determine which approach is best. In evaluating positive and negative tweets, the Naïve Bayes algorithm performed better than Sentimental Analysis into which the total number of positive, negative, and neutral tweets is counted. The technique for sentimental analysis is as follows: Choose a string of characters and evaluate them as positive, negative, or neutral. Still, the Sentimental Analysis algorithm performed better than Naïve Bayes when evaluating the neutral tweets on different airlines.

3. Airline Recommendations Using a Hybrid and Location-Based Approach

The author here used the dataset generated by him, which contains the data of about 480 flights with validations from the Google API class. Saurabh Baulker (2017) implemented a recommender system for airlines by combing the famous algorithms of Content-Based Filtering and Collaborative Filtering; the specialty here is to recommend an airline to a user on a personalized level rather than just supplying a list of airlines from an unexpected source to destination as recommendations. As the ratings parameter used for evaluating collaborative filtering between users to recommend flights was not possible, as it is challenging to give ratings to an airline merely relying on a single parameter, the latitude and longitude of the source and destination locations were used to implement Collaborative Filtering. The feasibility of evaluating airlines relying on a specific characteristic is problematic in the scenario of airlines since the many users will concentrate on various metrics that the airline sector utilizes, such as customer satisfaction, level of service, airline reliability, return on capital, and so forth. Therefore, collaborative filtering algorithms would need to be meshed up to provide a single parameter for producing ratings, which could then be used for filtering. Saurabh Baulker (2017) used the data provided by the user's profile to recommend the airlines to the user based on content-based filtering. For different combinations of origins and destinations of airlines like Vancouver-New York City, Los Angeles-New York City, and Pune-Delhi, the precision value was high, which means that the flights retrieved were also relevant.

4. **Flight Recommendation System based on user feedback, weighting technique, and context-aware recommendation system**

To assist users of service-oriented environments in discovering and choosing the most appropriate airline services from a vast number of accessible ones, a recommender system is implemented based on customers' preferences, weightage approach, and the context-aware system. Depending on prior purchases, this recommender system would give the customer preferred choice alternatives, basic info, and travel routes that best meet his tastes. The customer's tastes are learned through either direct or indirect input. The context-aware recommendation system is used to recommend airlines to a user because the context-aware recommender system considers the user's location, time, and extra info while making recommendations and making accurate flight recommendations to a user.

There are two essential mechanisms in this work: Preference Processor and Weight Processor, and in the preference processor, the overall counts of the choices are computed. It defines the preference characteristics based on the first element in the result set. Weight Processor's job is that the weights are fetched and progressively updated each time the user makes a decent purchase. The upgrading method entails a fractional weight increase for the selections made throughout the purchase. Non-selected attributes' weights are lowered at the very same time. This gives the system a significant advantage. The recommendation processor evaluates the results obtained from the preference and weight processor, and then the 'scores' are calculated for every item in the result set. The scores obtained after computation are ranked according to the sorting process, thus creating the final recommendation set, and then these results are updated in the customer's profile.

METHODOLOGY

We recommended airline names based on the user's choices. First, we take input from the user about their information like their nation, age group, feature preferences, and cabin preference. Using an API created by Stream lit, we use user input to recommend the airline. We recommend the best-suited airline based on the ratings using knowledge-based Recommendation (it makes recommendations based not on a user's rating history but on specific queries made by the user) on the user's input. The top five airlines are our initial recommendations. We used the Content-Based

Recommendation System, Collaborating Filtering Recommendations System, and Hybrid Recommendation System to find out more about similar airlines.

Content Based Recommendation System

In this case, we utilized K-Nearest Neighbor (which stores all the available cases and classifies the new data or case based on a similarity measure); here, it finds similar airlines based on similarity to the given airline. We chose five neighbors who are most like the content(features) provided by an airline in the Knowledge-Based Recommendation system (An Algorithm used by IMBD to recommend movies). We recommended similar airlines to the airline we advised based on user desire.

Collaborative Filtering Recommendation System

Another technique suggested similar airlines based on the user's preferences. First, we established a pivot table with User Id and Airline Name based on the User's Overall Rating of the airline. Then, using Singular Value Decomposition (SVD), we lowered the dimensionality of the data(matrix). SVD is a matrix factorization technique that decomposes any matrix into three familiar matrices. Next, we computed the correlation between the user's favorite airline and the preferences of other users. Then, we recommended the top ten percent that closely associated airlines with the user. The Collaborative System considers each user's preferences, finds a correlation with them, and suggests the best airline based on the correlation.

Hybrid Recommendation System

The Hybrid Recommendation system combines content-based and collaborative filtering recommendations. We used the SVD model in the hybrid to predict airline ratings. First, we used a content-based technique to locate similar users. Then we applied a collaborative approach to estimate a particular user's rating for other more connected airlines according to the other user's choice using SVD. A Hybrid Recommendation System advised the airlines based on the highest rated (Prediction) Airline.

MACHINE LEARNING OUTCOMES

1. **KNN classifier**

KNN is a perfect go-to model and a very good baseline for recommender system development. The k-Nearest Neighbors (KNN) method is a simple, supervised machine learning technique that may address classification and regression issues. Data is examined with an appropriate selection of k, reducing the number of mistakes. KNN stores all available cases and classifies new ones based on similarity or distance. The most frequent method for calculating distance is the Euclidean distance.

Figure 1. Classification Report of KNN Classifier

	precision	recall	f1-score	support
0	0.39	0.29	0.33	2725
1	0.60	0.70	0.65	4096
accuracy			0.54	6821
macro avg	0.49	0.50	0.49	6821
weighted avg	0.52	0.54	0.52	6821

2. **Random Forest**

Random forest is most accurate ensemble classifier and works efficiently on huge dataset. It can effectively predict the missing data accurately, even in situations where large portions of data are missing and without pre-processing. It combines bagging and random feature selection. Random forest contains decision trees that are combined individual learners.

Figure 2. Classification Report of Random Forest Classifier

	precision	recall	f1-score	support
0	0.93	0.94	0.93	2725
1	0.96	0.95	0.96	4096
accuracy			0.95	6821
macro avg	0.94	0.94	0.94	6821
weighted avg	0.95	0.95	0.95	6821

3. Gaussian Naïve Bayes

It is a Naive Bayes algorithm that assumes all the features to follow a normal distribution. We selected naïve bayes because it allows understanding and justifying the model's predictions more simply. It is a classification algorithm used for building models with massive datasets. It is more widely used in building a recommendation system for collaborative filtering to predict whether the user will like it.

Figure 3. Classification Report of Gaussian Naïve Bayes Classifier

	precision	recall	f1-score	support
0	0.92	0.94	0.93	2725
1	0.96	0.95	0.95	4096
accuracy			0.94	6821
macro avg	0.94	0.94	0.94	6821
weighted avg	0.94	0.94	0.94	6821

Table 1. Accuracies of Machine Learning Models

Model	Accuracy
1. KNN	54%
2. Random Forest	94.6 %
3. Gaussian Naïve Bayes	94.1 %

DATASET DESCRIPTION

Our dataset originates from Kaggle (https://www.kaggle.com/datasets/efehandanisman/skytrax-airline-reviews). It is widely used in recommender systems. The sample database schema is as follows:

Table 2. Dataset Variables

Column Name	**Type**	**Description**
Id	Integer	unique id for each airline name
airline name	String	age of each patient
Overall_rating	Integer	Overall point given to the trip between 1 to 10
author	String	Author of the trip
author_country	String	Author Country
cabin_flown	String	Purpose of the flight of the passengers (Personal Travel, Business Travel, Economy)

seat_comfort_rating	Integer	Satisfaction level of Seat comfort
cabin_staff_rating	Integer	Satisfaction level of Cabin Staff
food_beverages_rating	Integer	Satisfaction level of Food and drink
inflight_entertainment_rat-ing	Integer	Satisfaction level of inflight entertainment
Age	Integer	The actual age of the passengers

CONCLUSION

We evaluated several recommendation algorithms and their predictive capabilities. And we found many recommendation systems consider only rating and airline names to recommend the airline to the user. But here, we determine that other factors are also crucial in the airline recommendation system. User country, Age group, Feature preference, and Cabin choice are essential factors in recommending the most suitable airline for the user. Our study found that these factors affect the recommendation, and by that, users are getting more personalized and most suitable airlines. Our system is different from the traditional only rating-based system but more accurate. However, suppose a user is in another country (Not in his own country), in that condition, as we take the user country in the recommendation. In that case, our system cannot recommend an airline as accurately as to the same country user, which we will try to resolve in the future. Moreover, our system can be used to recommend airlines and movies, books, trains, and tourist destinations.

REFERENCES

Khaturia, D., Saxena, A., Basha, S. M., Iyengar, N., & Caytiles, R. D. (2018b). A Comparative study on Airline Recommendation System Using Sentimental Analysis on Customer Tweets. *International Journal of Advanced Science and Technology*, *111*, 107–114. https://doi.org/10.14257/ijast.2018.111.10

Airline recommendations using a hybrid and location-based approach. (2017c, June 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8250610

Tuteja., M. (2016, May 2). *Flight Recommendation System based on user feedback, weighting technique, and context-aware recommendation system | International Journal of Engineering and Computer Science*. International Journal of Engineering and Computer Science. http://www.ijecs.in/index.php/ijecs/article/view/2479

Cao, J., Xu, Y., Ou, H., Tan, Y., & Xiao, Q. (2018). PFS: A Personalized Flight Recommendation Service via Cross-Domain Triadic Factorization. *2018 IEEE International Conference on Web Services (ICWS)*. https://doi.org/10.1109/icws.2018.00039

Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. (2005, June 1). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/1423975

CHEN, L., HSU, F., CHEN, M., & HSU, Y. (2008). Developing recommender systems with the consideration of product profitability for sellers. *Information Sciences*, *178*(4), 1032–1048. https://doi.org/10.1016/j.ins.2007.09.027

Ziegler, C. N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. *Proceedings of the 14th International Conference on World Wide Web - WWW '05*. https://doi.org/10.1145/1060745.1060754

Indira, Dnvsls., (2016b, June 1). *Conglomeration of Instance Filtering's k- Nearest Neighborhood and Collaborative Filtering's Item-based Recommendation on Airline Dataset System using Map-Reduce and Mahout*. ResearchGate. https://www.researchgate.net/publication/344948259_Conglomeration_of_Instance_Filtering%27s_k-_Nearest_Neighborhood_and_Collaborative_Filtering%27s_Item_based_Recommendation_on_Airline_Dataset_System_using_Map-Reduce_and_Mahout