International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-1, May 2020

A Machine Learning Model for Recommending Restaurants based on User Ratings

B. Muni Lavanya, K.Kalyan Kumar, H.Shagufta Kayanath, D.Pushpalatha Bai

Abstract: ***However, oftentimes people just search a restaurant by using word "restaurant", while the word "restaurant" means differently to different individuals. For an Asian, it can mean a "Chinese restaurant" or "Thai restaurant". How to correctly interpret search requests based on people's preference is a challenge. Building a machine-learning model based on activity history of a registered user can solve this problem. The activity histories used by this research are reviews and ratings from users. This project introduces a data processing pipeline, which uses reviews from registered users to generate a machine-learning model for each registered user. This project also defines an architecture, which uses the generated machine-learning models to support real-time personalized recommendations for restaurant searching and type of foods good at those recommended restaurants. Finally, this project aims to develop a good machine learning model, different collaborative filtering methodologies are considered to predict restaurants using user ratings. Slope One, k-Nearest Neighbors algorithm and multiclass SVM classification are some of the collaborating methodologies are going to consider in this project.***

Keywords**:** ***Machine Learning, Regression, Training, predict, Accuracy.***

I. INTRODUCTION

Machine Learning is a analytical study of algorithms and statistical models that the systems used to perform a specific task without using any explicit instructions. It is very closely related to the computational statistics, which focuses on making predictions using computers. Machine Learning algorithms build a model based on sample data called training data. These are used in the wide range of applications.

According to a Search Engine Land survey conducted in 2014, 88% of the consumers determine the quality of local business based on online reviews and 72% of the consumers trust local business more because of the positive reviews over the business. What is a local business? It can be a medical practice, a restaurant, a gym, etc,. The result of this survey shows the important role of reviews in deciding to choose a local business.

Among various types of business, restaurant is the most searched category by users in Yelp. However, oftentimes people just search a restaurant by using word "restaurant", while the word "restaurant" means differently to different individuals. For an Asian, it can mean a "Chinese restaurant" or "Thai restaurant". How to correctly interpret search requests based on people's preference is a challenge. Building a machine-learning model based on activity history of a registered user can solve this problem. The activity histories used by this research are reviews and ratings from users.

II. LITERATURE SURVEY

In all the existing system models there are some drawbacks in those systems. So, we can come with an idea to reduce those drawbacks in the Recommender System.

[1]In Hotel recommendation based on surrounding environments paper, it explains about the restaurant's surrounding environment and user preferences are calculated based on their reviews over the restaurants. Then it calculates the similarities among the environments of each restaurant and user's preferences. Finally, it recommends top-k restaurants.

[2] In Location Based Personalized Restaurant Recommendation System for Mobile Environments paper, the recommendation is based on the user's foursquare data. The recommendation is only based on the foursquare data of users. In this system the recommendation is based on the user location and user history.

[3] In a restaurant recommender system based on user preference and location in mobile environment paper, the recommendation is based on the user history or user profile i.e, user visited restaurants based on the location. Baidu map cloud service is used to implement the proposed recommender system. Different types of algorithms are used in our project to which model is recommending accurately. Then based on the rmse, the efficiency of the models can be defined. Then finally, which model has lowest rmse that model is finalized as the model.

[4] Location, Time, and Preference Aware Restaurant Recommendation Method paper says that the recommendation score is computed by offline and online. Offline recommendation score is computed based on user's visiting trends, discovering preferences and restaurants popularity. Online recommendation score is computed based on the restaurant's distance.

Manuscript received on April 02, 2020.
Revised Manuscript received on April 15, 2020.
Manuscript published on May 30, 2020.
* Correspondence Author
B. Muni Lavanya, Assistant Professor (Adhoc), Department. of Computer Science and Engineering, JNTUACEP. Specialized in Computer Networks and Cloud Computing.
K. Kalyan Kumar, Student, Department Of Computer Science and Engineering, IV B.Tech
H. Shagufta Kayanath, Student, Department Of Computer Science and Engineering, IV B.Tech
D.Pushpalatha Bai, Student, Department Of Computer Science Engineering, IV B.Tech

III. COLLABORATIVE FILTERING ALGORITHMS

Collaborative Filtering is one of the method used for the purpose of recommender system. Collaborative Filtering has two sensed, they are narrow one and more general one.

In the narrow sense, this is a method making automatic predictions for the interests of users by collecting the user preferences. For example, collaborative filtering method for the selection of the television could make predictions about the which television show a user should like given a partial list of that user's tastes.

In the more general sense, collaborative filtering is the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. The applications of the collaborative filtering models involve the large amounts of datasets. There methods can be applied to the different types of data like sensing and monitoring data and financial data etc.

Content-based recommendation

Content-based approach is based on a user's behavior. For example, this approach might use
historical information, such the ratings given by the user to the movies he or she previously
watched. If a user gives high ratings to most of the action movies he watches, then content based
_ltering can use this history to identify and recommend movies with similar content
(action). This content can be manually de_ned or automatically extracted based on other
similarity methods.

The models described in the literature survey shows the data processing pipelines, which uses reviews and ratings from registered users to generate a machine-learning model for each registered user. Those also defines an architecture, which uses the generated machine-learning models to support real-time personalized recommendations for restaurant searching and type of foods at those recommended restaurants.

The Existing system is only based on the KNN algorithm which is used in the case of small amount of data. But when comes to the analysis of larger amounts of data other algorithms are used.

IV. PROPOSED SYSTEM

In this Proposed system, it aims to develop a good machine learning model, different collaborative filtering methodologies are considered to predict restaurants using cuisine ratings. Different Algorithms are used, those are Surprising Models, XGBoost and SVD Algorithm etc. And finally, efficiency of the all the models are compared to decide which model is best.

V. METHODOLOGY

Data Set :

In this system we have used the Yelp Data Set, which has divided the dataset into four parts. They are User Data, Restaurants Data, Restaurants Data and Cuisines Data.

Fig-1: Data Set Structure

Data used to train the model

UserID	PlaceId	Rating	Food Rating	Service Rating
U1077	135085	2	2	2
U1068	135104	1	1	2
U1067	132584	2	2	2
U1103	132663	1	0	2
U1070	132608	2	2	1

In the proposed system, the prediction of restaurants is done by the different models using the different algorithms such as Surprise Baseline Models, XGBoost Predictor, SVD algorithms and KNN Baseline Models etc. The Data set contains the information about the Restaurants and User Ratings. The Restaurant Reccomendation is maily based on the user ratings and user profiles and user history. It includes the location preference in the further developments.

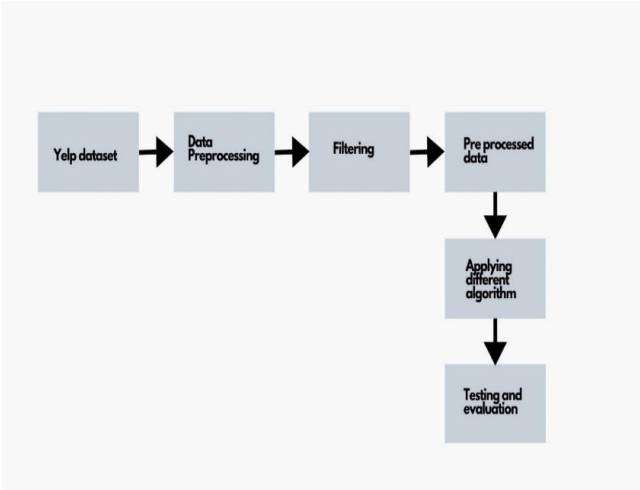

Fig-2: System Architecture

The Data set is collected from the Yelp.com site.

A. Pre-processing

The dataset contains all the information which the learning model is supposed to learn for making the right predictions.

The raw data might have a number of different values of each attribute which might lead to incorrect results. Therefore the information must pre-process by the learning process.
The pre-process includes the following:

The data set contain a lot of missing values. These missing values results in improper learning,these can be handled by following ways:

1)**Filling the missing values**: The missing values in the dataset must be filled with the most occurring data values in the data set for the attribute.

2)**Removal of Data instances**: The instances which contain missing values for any attribute must be removed. By this, we removed the unreliable data from the data set. So, by this technique, the number of data instances must be reduced. Hence this method can be implemented for large datasets.

3) **Splitting the Dataset into Training set and Test Set** : The instances which contain missing values for any attribute must be removed. By this, we removed the unreliable data from the data set. So, by this technique, the number of data instances must be reduced. Hence this method can be implemented for large datasets.

Converting to numerical values:

The attributes that contain string need to convert into numerical values. In the dataset the UserIDs are in the format of Strings. We need to Convert them into Numerical values for the easy calculations of the restaurant IDs.

Data Cleaning:

In the dataset, there can be some attributes which don't have an effect on the results. Such attributes need to be identified and can be removed from the dataset to simplify the dataset. We will take only the required data, that is taken as the input to the Machine Learning Models that is taken from the filtering step.

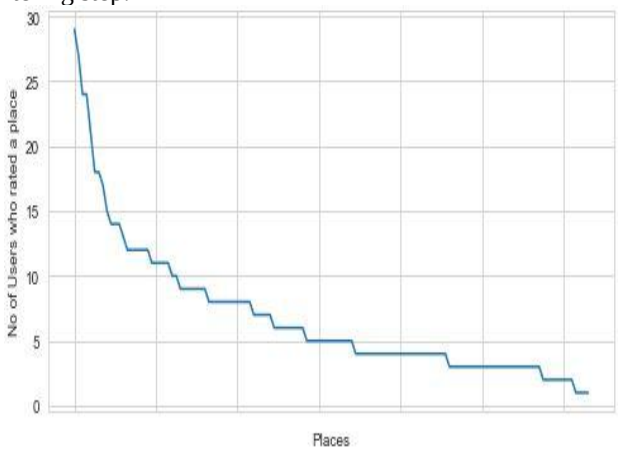

Fig-4: Ratings per Places

B. Implementation – Prediction:

Preprocessed Data

The Preprocessed data is taken from the data preprocessing step and finally the statics of the data is shown below :
Data shown below is ratings to the number of Ratings

Fig-3: Distribution of Ratings

In the prediction process, we will take the similarities among the user ratings given to the restaurants that the users has visited.

C. Algorithms Used:

In our Restaurant recommender system we are using the following algorithms and finally compare the efficiency of all the algorithms to decide the best model :

- XGBoost
- SVD Algorithm
- Surprise Baseline Model
- Surprise KNN baseline Model

XGBoost is an algorithm that has recently been dominating applied Machine Learning and Kaggle competitions for structured or tabular data. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

Singular Value Decomposition (SVD) provides another way to factorize a matrix into singular values. Mainly SVD algorithm used to matrix operations and data reductions.

Surprise Baseline Model is an algorithm mainly used for predictions. What the algorithm going to done is predicts the random rating based on distribution of the training set.

Surprise KNN Baseline Model is an algorithm, in which the actual number of neighbours that are aggregated to compute an estimation is necessarily less then or equal to k.

VI. RESULT AND ANALYSIS

In figure 3, we have taken the data distribution over the user ratings and and the number of users ratings. Using this data we can easily train the machine to predict the similar restaurants based on the ratings. And also from figure 4, you will find the number of ratings given by the uses to the different restaurants or places. Then we will apply the similarity matrix to the filtered data and we finally get the result. Here we can customize the maximum number of similar restaurants to be given as output. In this, there are top 10 restaurants those are similar to the user searched restaurant.

For example, here we will take the 135041 as the restaurant ID and then the model will suggest the top 10 restaurants similar to the user searched restaurant.

```
Name of the restaurant -------> Luna Café
Number of ratings by users for restaurant
Luna Cafe is 14
Number of similar places to Luna Cafe is
38
```

Then finally the top 10 recommended restaurants are:

PlaceID	Name
135051	Restaurante Versalles
135106	El Rincón de San Francisco
135039	Restaurant de Mariscos de Picon
135052	La Cantina Restaurante
135060	Restaurante Marisco Sam
135108	Potzocalli
135054	Restaurante y Pescaderia Tampico
135032	Cafeteria y Restaurant El Pacifico
135057	El Herradero Restaurante and Bar
132885	Hamburguesas saul

On building the model with different algorithms will gives the different types of recommendations. Then the best model can be decided based on the efficiency. Efficiency is based on the RMSE and MAPE. The following table compares the efficiency of the different algorithms:

Algorithm	RMSE
svd	0.4948488837887633
knn_bsl_u	0.4965311998141357
knn_bsl_m	0.4992930966551378
bsl_algo	0.4972010487454593
xgb_all_models	0.5011586550903637
svdpp	0.5018114296611016
xgb_final	0.5457415078379207
xgb_knn_bsl	0.5458901322913035
first_algo	0.5464312409409059
xgb_bsl	0.5502140506348814

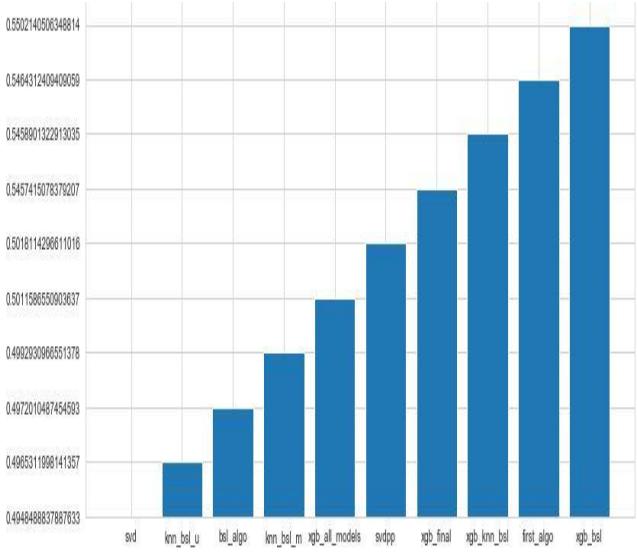

Fig-5. Comparing the different models accuracy in predicting the Restaurants

Then, the model trained with SVD algorithm is very efficient and perfect to recommend the restaurants based on the user history.

VII. CONCLUSION

As major half of the world is becoming socially active we made efforts to generate recommendations using ratings from Online booking Food Systems like Zomato, Swigge, user's location and behavior of user while using the application. We have used Collaborative filtering which makes our system more effective and accurate. We have compared different models using the different algorithms to reach the highest accuracy. High accuracy is nothing but the perfect recommendation of the restaurants based on the user profiles and history. Thus, we can conclude that the proposed restaurant recommender system can effectively utilize user's preference and available ratings of restaurant to recommend the personalized and suitable restaurants for different users. This recommender system adopted a user preference model by using the features of user's visited restaurants, and also utilized user preferences and restaurants to dynamically generate the recommendation results.

REFERENCES

1. In Hotel recommendation based on surrounding environments paper they have evaluated each hotel's surrounding environments with the help of POIs databases, calculated the preferences of the users using their reviews of hotels that they stayed, calculated the similarity between the environment of each hotel and user's preferences and select the top-k hotel to recommend to the user.
2. In Location Based Personalized Restaurant Recommendation System for Mobile Environments paper where did recommendation by using user's foursquare data and by locating the user's location. The system recommends the restaurants to the new user by using his foursquare data.
3. Jun Zeng; Feng Li; Haiyang Liu; Junhao Wen; Sachio Hirokawa 2016 5th IIAI International conference on Advanced Applied Informatics (IIAI-AAI) "A Restaurant Recommender System Based on User Preference and Location in Mobile Environment".
4. Md. Ahsan Habib; Md. Abdur Rakib; Mohammad Abul Hasan 2016 19th International Conference on Computer and Information Technology (ICCIT). "Location, time, and preference aware restaurant recommendation method"

AUTHORS PROFILE

B. Muni Lavanya, Assistant Professor (Adhoc), Department. of Computer Science and Engineering, JNTUACEP. Specialized in Computer Networks and Cloud Computing.

K. Kalyan Kumar, Student, Department Of Computer Science and Engineering, IV B.Tech

H. Shagufta Kayanath, Student, Department Of Computer Science and Engineering, IV B.Tech

D.Pushpalatha Bai, Student, Department Of Computer Science Engineering, IV B.Tech