TravelBot: Utilizing social media dialogue for Travel Recommendation

Nandarani Kadam[1], Sarika Solanke[2]

[1,2]*Computer Science Engineering, Deogiri Institute of Engineering & Management Science, India*

Abstract - **Use of social media in tourism is increasing rapidly. Due to increasing Internet usage, many businesses now use online platforms to handle customer inquiries, and many of them turn to chatbots for improving their customer service. Travelbot is a chatbot; a computer application that interacts with users using natural language in a similar way to imitate a human travel agent. Travelbot is using twitter data for recommending countries to users for travel. The travelbot uses tweepy api and perform filtering on data to get all travel related tweets. When travelbot recommended countries to travel realized that there are number of features available like safety, quality of life, healthcare etc. So all these features are important? To solve this question need to use classification algorithm. First Decision tree algorithm is used with these datasets, on the values of feature importance got accuracy score 48.39%. Then Random Forest Algorithm is used with these datasets, on the values of feature importance got accuracy score 78.65%. According to these values we can see the accuracy of random forest is increased 30% as compare to decision tree algorithm.**

Index Terms - **Artificial Intelligence, Chatbot, Machine Learning, Natural language processing, Travel, Twitter.**

I. INTRODUCTION

A chatbot, also known as a conversational agent, is computer software capable of taking a natural language input and providing a conversational output in real time [1]. The human and chatbot communicate using GUI i.e. Graphical user interface and which is based on HCI i.e. Human Computer Interaction principles. The idea of an automated computer program engaging in human interactions was first theorized by Alan Turing in 1950. Shortly after, automated computer programs, referred to as "bots", were created to motivate human conversation. Chatbots are typically used in dialog systems for various purposes including customer service, request routing, or for information gathering. While some chatbot applications use extensive word-classification processes, Natural Language processors, and sophisticated AI, others simply scan for general keywords and generate responses using common phrases obtained from an associated library or database [2]. A study by Forrester (June 2017) predicts that 25% of today's jobs will most likely be impacted by AI technologies by 2019. Chatbots now exist in various messaging platforms, such as Facebook Messenger, Skype, and Kik, largely for customer service purposes.

As an emerging trend most travel companies provide online services. The rapid growth of online travel information increases the challenges for tourists also increases. The tourist who has to choose from a large number of available travel packages for satisfying their needs. The travel companies have to understand tourist preferences to increase their profit. So what travel companies do, they use intelligent travel services. This has been driven, in part, by passenger desire to use social media as a channel for improved communication with transport providers; while transport operators have also recognized the value to be gained from insights about the passenger experience [2]. Conversations between passengers and operators are perhaps most frequently seen on Twitter (Passenger Focus and Abellio, 2012), an online social media platform that allows users to share short messages (originally 140 characters, though raised in November 2017 to 280 characters) and engage in dialogue via responses, re-Tweets, and shared hashtags.

The travelbot is recommending you countries to travel using important feature. So that by using graphs the output results can be analyzed. If user does not have any information about countries, then also the bot can recommend you. Then follow the steps which are mentioned below.

- User interact with bot (Greetings).

- User reply travel suggestions?
- Bot give response some options.
- User selects any option from the list.
- Bot give the result of that option.

First, we collect international data from kaggle and data. world site. After collecting country data, we organize in a proper format. We perform data cleaning steps and clean all data. After cleaning we normalize our data. When we collect data from different source the data is not in a proper format, we have to organize it. So we remove the countries don't have much data in it. We replace NaN value with mode.

Now we use random forest algorithm to extract the best features from data. Now we collect data from twitter using tweepy API. When we extract data, data is not in a format. So we save all data in a csvfile then perform operations on that data. First, we remove RT (retweet), if there are 1000 tweets then only 10-20 tweets are original. So we have to remove retweet from our data. After that we normalize it by using some travel related keywords. The data from twitter related to travel contain stop words so we remove that also. Now we get a wordcloud which is very important.

II. OBJECTIVE

A chatbot is a service that people interact with via a chat interface. You can ask questions by typing in the same way you would ask a person. The chatbot will usually respond in a conversational style, and it may carryout actions in response to your conversation (for example, order something for you) [3]. It answers your question, rather than directing you to a website. The Chatbot uses machine learning and artificial intelligence to provide the best response. We will call these AI-powered chatbots. It understands language, as well as commands. It has the ability to constantly learn from user interactions to become better at predicting their needs. The data that traditional travel agent has can be limited or outdated, this makes agents not so efficient. Chatbot can chat in a similar way a staff member would with a person. Chatbot can store and categorize the information it receives from each interaction [4]. It can assess information to identify which information is of no value and which is not. In addition to that, when a customer calls a travel agent or a travel company after some months, most of them forgot the history and the interests of the caller.

Chatbot knows where to store that information, so it can access it again in the future.

III. LITERATURE SURVEY

Qi Liu et at "A Cocktail Approach for Travel Package Recommendation" (Feb 2014) conclude that Cocktail Model is the combination of all Models. As compared to all other model the cocktail model can give best result. The cocktail model can work on the user profiles and travel logs [5]. The TAST Model is used for detecting tourist interest and topics the output of this model can give input to the Cocktail Model. The Collaborative filtering can be performing on those packages and remove unwanted packages. Shuhui Jiang, et al concluded" Personalized Travel Sequence Recommendation on Multi- Source Big Social Media" (Mar 2016) in which by understanding package model from huge multi-source social media and community shared pictures, a personalized travel sequence Recommendation system [6]. This system automatically mine users' interest and routes travel topical preferences such as the topical interest, cost, time and season. It recommends not only POIs but also sequence of travel. This system recommended not only POIs but also travel sequence, considering bot user's travel preferences at the same time.

Paul Gault et al ": Utilizing social media dialogue to provide journey disruption alerts" (Nov 2019) conclude that the potential for utilizing social media to provide personalized information to public transport passengers, drawing from lessons learned from related studies. The Tweeting Travel study developed an understanding of the types of tweets on social media between passengers and a travel advice system and then used this to design of the TravelBot system [1, 7].

There are many websites which give result for travel guidance, but user get confused which travel site is good. We observe that many sites are there to give result, but we cannot say which company is good for user. To select the travel package user searches 100 of sites. To ease to find the good result here we are developing a Chatbot. There are many travel agents available, but travel agents are a human being they have time limitation. Travel agents also cannot available 24/7 [8]. No Travel agent in this world can "know it all". Chatbot is an AI so it can learn from user

and save the information for later use. We are using twitter data to train the Chatbot.

Now a day's number of people is active on social media. They post whatever they do. So, by using their tweets we can give result to the user. The tweet and re-tweet contain different type of information, but we are using algorithm and filter the data. The data contain personal information, or some sensitive information. We need only travel related information, so we filter this real time data and give to our Chatbot [9]. There is too much data available in twitter so when users ask for something what Chatbot do? Chatbot take all information from twitter and give them (priority) confidence value. Then Chatbot look confidence values for each statement give back result to user, which value is having best match for user's question.

The program selects the closest matching response by searching for the closest matching known statement that matches the input; it then chooses a response from the selection of known responses to that statement. An untrained instance of Chatbot starts off with no knowledge of how to communicate. Each time a user enters a statement, the library saves the text that they entered and the text that the statement was in response [10]. As Chatbot receives more input the number of responses that it can reply and the accuracy of each response in relation to the input statement increase. Chatbot uses logic adapters. Logic adapters determine the logic for how Chatbot selects a response to a given input statement. The logic adapter that your bot uses can be specified by setting the logic_adapters parameter to the import path of the logic adapter you want to use. It is possible to enter any number of logic adapters for your bot to use. If multiple adapters are used, then the bot will return the response with the highest calculated confidence value. If multiple adapters return the same confidence, then the adapter that is entered into the list first will take priority. A confidence value and the selected response statement should be returned. The confidence value represents a rating of how accurate the logic adapter expects the selected response to be. Confidence scores are used to select the best response from multiple logic adapters. The confidence value should be a number between 0 and 1 where 0 is the lowest confidence level and 1 is the highest.

IV. PROPOSED SYSTEM

As shown in figure system start execution, when user visit the web page.
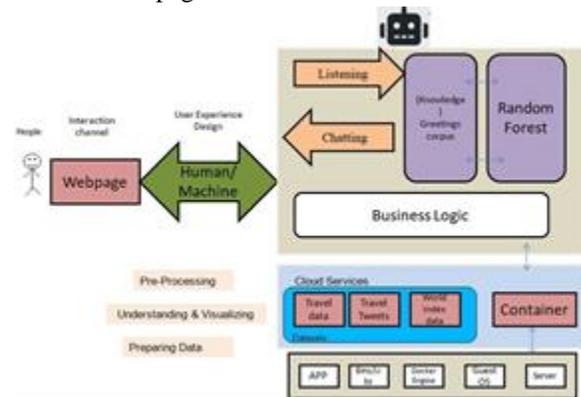
Fig. 1 Proposed System

1. The bot greet user.
2. User type keyword travel then bot gives the response.
3. After this listening ant chatting knowledge block, which contain the greeting corpus.
4. Here we are using social media dialogue as datasets.
5. When user ask for travel, then using machine learning algorithm we get the important feature.
6. Only that feature bot suggests to the user.
7. After selecting feature by user bot give countries to travel.

Algorithm
Random Forest Classifier
Input: Dataset with all features
Output: Feature Importance and accuracy score
Step 1: Pick at random k data points from the training set.
Step 2: Build a decision tree associated to these k data points.
Step 3: Choose the number N of trees you want to build and repeat steps 1 and 2.
Step 4: For a new data point, make each one of your N-tree trees predict the value of y for the data point in question and assign the new data point to the average across all of the predicted y values.

V. EXPERIMENTAL RESULTS

5.1. Dataset Details
In the experiments, the following three datasets are used.

• Kaggle

This dataset is a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment.

Link to Dataset: https://www.kaggle.com/search?q=travel+data

• Numbeo

This dataset collected by Mladen et al. Available data at Numbeo.com are not influenced by any governmental organization. Numbeo is the world's largest cost of living database. Numbeo is also a crowd-sourced global database of quality-of-life informations including housing indicators, perceived crime rates, and quality of healthcare, among many other statistics.

Link to Dataset: https://www.numbeo.com/quality-of-life/rankings_by_country.jsp

• Data.world

Our cloud-native data catalog maps your process, distributed data to familiar and consistent business concepts, creating a unified body of knowledge anyone can find, understand, and use.

Link to Dataset: https://data.world/anish-philip3/tourism

• Tweepy

We are using twitters tweepy api for getting real time tweets of users. Then we filter all data and get only travel related country names and rank all countries.

Link to Dataset: https://twitter.com

5.2 Performance Metrics

Precision is the ratio between the True Positives and all the Positives [11]. For our problem statement, that would be the feature importance that we correctly identify important for travel out of all the features. Mathematically:

Precision = True Positive (TP) / (True Positive (TP) + False Positive (FP))

The recall is the measure of our model correctly identifying True Positives. Thus, for all the features who actually important, recall tells us how many we correctly identified as having important to travel. Mathematically:

Recall = True Positive (TP) / (True Positive (TP)

+ False Negative (FN))

F-score is the measure of accuracy test and it considers the both precision P and recall R of the test in order to calculate the score. The general formula for F-score measure is given in the equation.

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

Here we are calculating feature importance values using feature_importances.

5.3 Performance Analysis of Proposed work

Accuracy score can show exactly the accuracy of the model. Here we are calculating accuracy score using decision tree and random forest algorithm. After calculating values, we conclude that random forest work better as compared to decision tree algorithm. Accuracy score values are shown in below table.

Datasets	Decision Tree	Random Forest
Travel Tweet		
Travel Data	48.39%	78.65%
World Index Data		

Table 5.1 Accuracy Score

We plot graph using these values shown below.
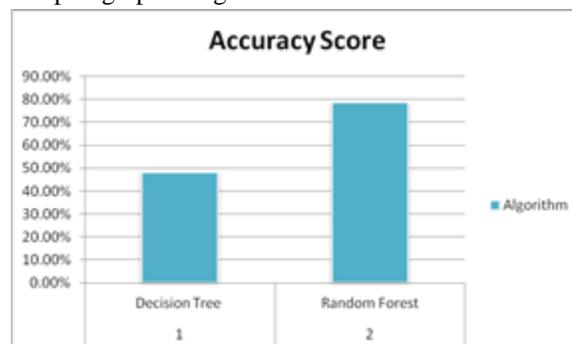
Fig. 5.1 Accuracy Score Graph

Above Details shows the comparison of Decision Tree algorithm accuracy score with Random Forest Algorithm. We can see that Random Forest made the most accurate predictions thus far with an improvement of 30% from the last model!

Feature Importance

In table 5.2 we calculate feature importance values of each feature. So from above values we cannot conclude anything. Here we plot a graph using above values. Now we can see that from 12 features only 7 features are important. When people travel then they consider is that country is safe, quality of life is good

and is country is trending? So we can say that using random forest algorithm we get 7 features and only those features we are choosing to recommend travel to the user. We train our bot using only important features. We calculate the accuracy score and compare it with decision tree algorithm. Here can say that as compare to decision tree random forest work better for our project.

No.	Decision Tree dt.feature_	Random Forest rfc.feature_
1	0.04240125	0.08730896
2	0.24048171	0.08293712
3	0.02354583	0.08911921
4	0.12648562	0.0883668
5	0.05783882	0.08598398
6	0.11975937	0.08860118
7	0.10006978	0.08758722
8	0.05519993	0.09733622
9	0.1147127	0.0653466
10	0.06388797	0.08782316
11	0.0438441	0.06782631
12	0.01177292	0.07176324

Table 5.2 Feature Importance Values
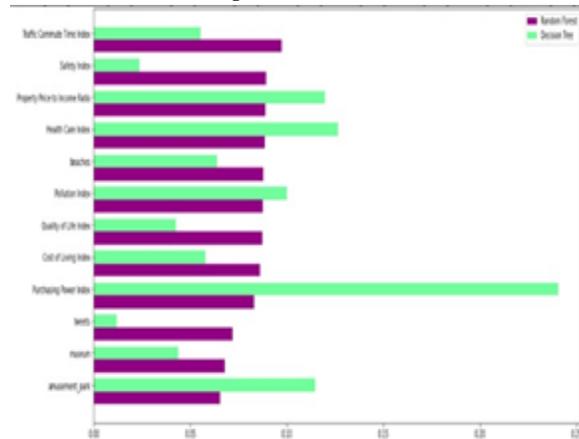
Fig. 5.2 Graph of Important Features

VI. CONCLUSION

Travelbot is using twitter data to recommend the travel countries to user. The travelbot uses tweepy api and perform filtering on data, after that bot get all travel related tweets. These tweets dataset and travel dataset are used for implementation. After combining these datasets found number of features available to explore. Random Forest algorithm and Decision Tree algorithm both are the classification algorithms. After getting the values of feature importance for both the algorithms according to that we conclude which feature is important for recommending travel. So we can say that the features we got by using random forest algorithm are useful for travel. Based on the mentioned research work in this paper which has been diverted with the base paper in many ways, following conclusions can be drawn:

- The research work proposes to use of travelbot is less hectic than others, performance enhancement because of twitter data.
- Moreover, the bot gives you quality of life, safety, pollution index for countries which is also important features when you travel.

Limitations

Despite of encouraging results of travelbot, the work is limited in the following ways:

- Booking of tickets is not possible.
- It does not give you hotels and car rentals.

VII. FUTURE SCOPE

The proposed research work can be extended in many ways as follows:

- The bot can upgrade by using maps.
- This bot can use travel companies.
- This bot is only text based, can be voice based.
- Integrate it with Facebook.

REFERENCES

[1] R. Khan and A. Das, "Introduction to chatbots," in Build Better Chatbots, Berkeley, CA: Apress, 2018, pp. 1–11.
[2] A. Følstad and P. B. Brandtzæg, "Chatbots and the new world of HCI," Interactions. ACM.org, vol. 24, no. 4, pp. 38–42, Jun. 2017.
[3] A. Schlesinger, K. P. O'Hara, and A. S. Taylor, "Let's talk about race: Identity, chatbots, and AI," in Proc. the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), 2018, pp. 1–14.
[4] R. Dale, "The return of the chatbots," Natural Language Engineering, vol. 22, no. 5, pp. 811–817, vol. 10673 LNCS, pp. 377–392.

[5] "A Cocktail Approach for Travel Package Recommendation", Qi Liu, Enhong Chen, Hui Xiong, Yong Ge, Zhongmou Li, and Xiang Wu IEEE Transactions on Knowledge and Data Engineering, Vol. 26, No. 2, February 2014.

[6] Shuhui Jiang, XuemingQian, Tao Mei and Yun Fu, "Personalized Travel Sequence Recommendation on Multi- Source Big Social Media", in 2016. IEEE TRANSACTIONS ON BIG DATA, VOL. X, NO. X, SEPTEMBER 2016.

[7] Paul Gault, Caitlin D. Cottril, "TravelBot: Utilising social media dialogue to provide journey disruption alerts", Transportation Research Interdisciplinary Perspectives 3 (2019) 100062.

[8] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews", In ACM CSCW'94, pp. 175–186, 1994.

[9] P. B. Brandtzaeg and A. Følstad, "Why people use chatbots," in International Conference on Internet Science (INSCI 2017), 2017.

[10] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Application of dimensionality reduction in recommender systems-a case study, In ACM WebKDD Workshop, pp. 82–90, 2000.

[11] Herlocker, J. Konstan, L. Terveen, and J. Riedl, "Evaluating Collaborative Filtering Recommender Systems," ACM Trans. Information Systems, vol. 22, no. 1, pp. 5-53, 2004.

[12] Prissadang Suta, "An Overview of Machine Learning in Chatbots", International Journal of Mechanical Engineering and Robotics Research Vol. 9, No. 4, April 2020.