International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 10 | Oct 2022 www.irjet.net p-ISSN: 2395-0072

Tourist Destination Recommendation System using Cosine Similarity

Shashank Jagtap[1], Sayali Borate[2]

[1]Student, School of Science, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India,
[2]Student, Dept. of Computer Engineering, PES Modern College of Engineering, Pune, Maharashtra, India

---

Abstract - *Over the past decade, internet has become our go-to tool for information and even for recommendations. Internet is used to find places for vacations. However, with the vast amount of destination options and its information, a lot of time is wasted before a relevant tourist destination is ascertained. The Destination Recommendation System leverages on Data Analysis and Machine Learning in order to give cogent and fast recommendations. This paper describes an approach which offers generalized recommendations to every user using the Cosine Similarity algorithm. The dataset used exhibits a vast and distinct combination of tourist places. The Cosine Similarity algorithm predicts the most relevant tourist places using some important features of the dataset such as tourism category, minimum budget(per day) and the visa requirement.*

Keywords: ***machine learning, cosine similarity, recommendation system, tourist destination***

1. INTRODUCTION

The amount of information is increasing day by day. To traverse this information in order to find relevant information is difficult. To automate the process of traversing this enormous information, we use machine learning algorithms. They are of 3 main types[2]: 1. Supervised: learning a function that maps an input to an output based on example input-output pairs. 2. Unsupervised[11]: algorithms are left to their own devices to discover and present the interesting structure in the data without a supervisor. 3. Reinforcement: algorithms are concerned with how software agents ought to take actions in an environment in order to maximize some notion of cumulative reward. We use these machine learning algorithms to find patterns and similarities between different items of a dataset. A recommendation system is an application of machine learning that plays a vital role in providing relevant recommendations, be it a movie recommendation or a product recommendation. In our daily life we depend on recommendations provided by our friends and family or general surveys. Similarly, recommendation systems are tools used to provide logical and rational product recommendations to users that might interest them by using some algorithms.

The Destination Recommendation System filters through enormous data and provides a highly relevant and cogent recommendation based on vital parameters of the tourist dataset. It uses Cosine Similarity algorithm to provide fast and reliable recommendations.

Our goal is to minimize user effort and provide a fast and one stop solution for finding their dream vacation place.

1.1 Dataset Attributes:

- Destination: This specifies the tourist destination. It consists of all the possible tourist destinations and the output predicted is based on the input destination provided by the user.
- Category: This attribute describes the category of a particular tourist place. It specifies whether the place has a historical, cultural, architectural, commercial or any other significance.
- Minimum budget per day: It specifies the minimum daily budget. This variable is expressed in terms of dollars($).
- Best Months: This attribute gives the best time to visit a particular destination based on climatic conditions.
- State/ Country: It specifies the location of the destination.
- Continent: Tells about the continent of the destination.
- Language: This variable describes the native language of a particular destination.
- Visa: This attribute tells about the visa requirement for Indian passports. It is a boolean variable.

1.2 Phases of recommendation process:

There are three main phases[3] in the process of recommendation as follows:

1. Information collection phase : In this phase, relevant information about the users is collected in order to generate a user profile. The system needs to know as much as possible from the user in order to provide reasonable recommendations.

Recommendation systems rely on different types of input such as :

- Explicit Feedback: In this, the system asks the users to provide their ratings for items via the user interface. The quality and efficiency of the recommendation system relies on the user ratings.
- Implicit Feedback: The user's preferences are automatically inferred by monitoring different actions such as navigation history, time spent on web pages, links followed, etc. There is no user participation required to gather implicit feedback, as the explicit feedback.
- Hybrid Feedback: It is a combination of both-explicit and implicit feedback. It works by using an implicit data as a check on explicit rating or allowing the user to give explicit feedback only when he chooses to express explicit interest.

2. Learning Phase : This phase applies learning algorithms on the user's data which are obtained from the feedback in the information collection phase. In certain situations, the learning algorithms are the methods which are helpful in drawing out the patterns appropriate for application.

3. Recommendation Phase: The pattern obtained from the previous phase is analyzed in order to provide recommendations for given data. This recommendation can be made either directly based on the dataset collected in information collection phase or through the system's observed activities of the user.

2. RELATED WORK

Recommender systems take into account user preferences and then suggest personalized content.

Recommender systems can be designed using any of the following approaches[1]:

1. Content-based filtering
2. Collaborative filtering
3. Hybrid filtering

1. Content-based filtering[14]: Content-based recommenders are essentially a user-specific learning problem to quantify the user's utility (likes and dislikes, rating, etc.) based on item features. Content-based filtering provides recommendations based on previously used feedback. It does not require other users' data during recommendations to one user.

2. Collaborative filtering[13] : Collaborative filtering filters information by using the interactions and data collected by the system from other users. This technique uses the similarity index-based technique. This filter can filter out items that users like on the basis of ratings or reactions by similar users. There are 2 sub types-

    a. User - user based - this measures the similarity between target users and other users.

    b. Item - item based - this measures the similarity between the items that target the users rate or interact with and other terms.

3. Hybrid Filtering: Hybrid filtering is the combination of Content based filtering and Collaborative Filtering. It is used to eliminate the limitations of content-based and collaborative filtering and generate more accurate recommendations.

3. METHODOLOGY

The project aims to build a platform that will recommend tourist destinations to users, provides a detailed description of the recommended destination. The information provided cuts down the time spent in going through multiple websites in order to decide a specific destination to visit.

The model allows us to predict the best tourist destination based on user input. It uses a content based filtering algorithm, cosine similarity, to make these recommendations.

Following are the steps used for recommending destinations to the user

1. Data collection
2. Data preparation
3. Combining relevant features
4. Apply filtering algorithm
5. Provide recommendations

1. Data collection: The first step to build a destination recommendation system is getting the appropriate data. In this project, we designed the dataset. The dataset used contains

9 features namely, Index, Destination, Category, Min Budget(Per Day in $), Best Months, State/Country, Continent, Language and Visa.

2. Data Preparation : After data collection, preprocessing the data to handle corrupted or missing data is essential. This is done in the data preparation step.

3. Combining relevant features: In this step, only the features required to make recommendations will be combined into a single attribute. In destination recommendation system, 3 features(Category, Min Budget(Per day in $) and visa requirement) will be used to make predictions.

4. Apply a filtering algorithm: To recommend destinations, a filtering algorithm needs to be implemented in order to find the amount of similarity between different destinations available in the dataset. In this project, a content based filtering algorithm - cosine similarity is used to make recommendations.

5. Provide recommendations: Once cosine similarity for a particular user input has been applied, different data records in the dataset will get different values. Based on this, after sorting, the top 5 destinations will be recommended to the user.

The proposed model uses cosine similarity in order to recommend tourist destinations based on user input.

Cosine similarity[5][12] : Cosine similarity is a metric, helpful in determining, how similar the data objects are irrespective of their size. We can measure the similarity between two records using cosine similarity. The comparison is done by finding the dot product between the two identities. The formula to find the cosine similarity is,

where,

$$\cos(x, y) = \frac{x.y}{||x|| * ||y||} \qquad (1)$$

- x . y = product (dot) of the vectors 'x' and 'y'.
- ||x|| and ||y|| = length of the two vectors 'x' and 'y'.
- ||x|| * ||y|| = cross products of two vectors.

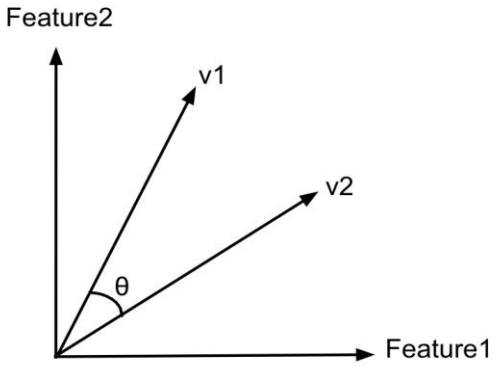

Fig-1: Cosine Similarity

As the above diagram shows, the angle between v1 and v2 is Θ. Greater the angle between the two vectors lesser is the similarity. It means if the angle between two vectors is large, they are very different from each other and if the angle between the two vectors is small, then the vectors are almost alike.

The cosine distance is calculated using the formula:

cosine distance = 1 - cosine similarity

Hence, when

- Θ = 0

  cos 0 = 1

  cosine distance = 1 - 1 = 0

  ∴ The two vectors are same
- Θ = 90

  cos 90 = 0

  cosine distance = 1 - 0 = 1

  ∴ The two vectors are very different
- Θ = 180

  cos 180 = -1

  cosine distance = 1 - (-1) = 2

  ∴ The two vectors are opposite to each other

4. RESULT

In this recommendation system, we have used Cosine Similarity and Content-based filtering to recommend a destination to the user. The code is written in python language and uses NumPy and pandas library.

The project uses cosine similarity in order to determine what destinations are to be recommended to the user. Cosine similarity measures the angle between two vectors and determines the amount of similarity.

In Tourist Destination Recommendation System, the user is first asked to enter a destination of his liking. Cosine similarity algorithm is implemented on the destination entered by the user, thus generating a cosine distance for every destination available in the dataset. Parameters- Minimum budget, visa requirement and category are used to determine the similarity. Once all destinations have been assigned a score, they are sorted in descending order. The top 5 destinations similar to the one entered by the user are then recommended to him. Additional information about the location, budget, visa requirements, best time to visit is also provided to the user.

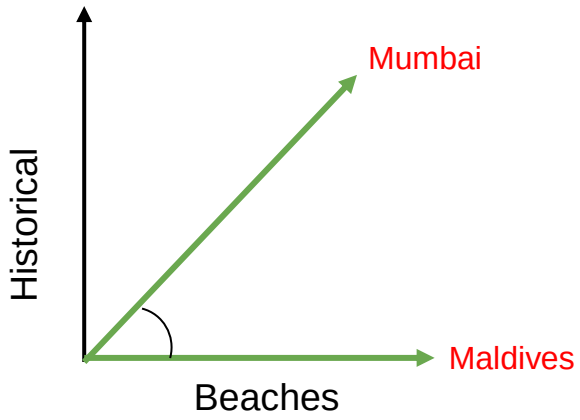

Fig-2: Cosine similarity depicting two different categories- Beaches and Historical

In this diagram, we have taken two categories - 'Beaches' on x-axis and 'Historical' on y-axis. When the user enters the destination as "Maldives", all destinations available in the dataset will be ranked. To measure the similarity, the angle between them is denoted by theta(Θ). Similarity ranges between 0 - 1. Mumbai lies in both categories (Historical, Beaches). Hence, when the destinations are ranked there is a high possibility of Mumbai getting recommended after places with beaches category are recommended; rather than a place which is only historical.

Output when the user enters "San Francisco" as his preferred destination is shown in Fig-3.

5. CONCLUSION

The proposed system recommends similar destinations based on the user's input which is a place of his liking. Every destination in the dataset is ranked with the help of the principle of Cosine Similarity and is then sorted before providing the recommendations. Currently, the recommendations are based on only three parameters; but the system can be improved using more parameters. New destinations can be added to the dataset to provide more recommendations.

```
Enter a destination that you liked in your past visits: San Francisco

Recommendations:
```

No.	Destination	Daily Budget	Best Months to Visit	State	Continent	Language	Visa
1	New York	238	Apr-June & Sep-Nov	New York	North America	English	Yes
2	Las Vegas	246	Mar-May & Sept-Nov	Nevada	North America	English	Yes
3	Austin	60	Sep-Nov	Texas	North America	English	Yes
4	Albuquerque	108	Sep-Nov	New Mexico	North America	English	Yes
5	Kiev	28	Apr-June	Ukraine	Europe	Ukranian	Yes

Fig-3 : Top five recommendations provided when the user enters "San Francisco" as his liked destination

6. REFERENCES

[1] Ramni Harbir Singh, Sargam Maurya, Tanisha Tripathi, Tushar Narula, Gaurav Srivastav, Movie Recommendation System using Cosine Similarity and KNN, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958

[2] Batta Mahesh, Machine Learning Algorithms - A Review, International Journal of Science and Research (IJSR)ISSN: 2319-7064

[3]F.O.Isinkaye, Y.O.Folajimi, B.A.Ojokoh, Recommendation systems: Principles, methods and Evaluation, Egyptian Informatics Journal, Volume 16, Issue 3, November 2015, Pages 261-273

[4] Prof Sachin Walunj, Yashwant Bhaidkar, Pranav Bhagwat, Priyanka Bhalere, Radhika Gujar, Tourist Place Recommendation System, IJARIIE-ISSN(O)-2395-4396

[5] Hieu V. Nguyen and Li Bai, Cosine Similarity Metric Learning for Face Verification

[6] Mallari Vijay Kumar, P.N.V.S. Pavan Kumar, A Study on Different Phases and Various Recommendation System Techniques, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-5C, February 2019

[7]Shubham Pawar, Pritesh Patne, Priya Ratanghayra, Simran Dadhich, Shree Jaswal, Movies Recommendation System using Cosine Similarity, International Journal of Innovative Science and Research Technology ISSN No:-2456-2165

[8] Sowmya D, Sayyed Johar, Ganavi M, Sankhya N Nayak, Analyzing Wine types and Quality using Machine Learning Techniques, International Journal of Engineering Applied Sciences and Technology, 2019, ISSN No. 2455-2143

[9] Pravinkumar Swamy, Sandeep Tiwari , Kunal Pawar Information Technology, Prof.Bharati Gondhalekar, Tourist Place Recommendation System, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181

[10] Prof. P. A. Manjare, Miss P. V. Ninawe, Miss M. L. Dabhire, Miss R. S. Bonde, Miss D. S. Charhate, Miss M. S. Gawande, Recommendation System Based on Tourist Attraction, International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395 -0056, p-ISSN: 2395-0072

[11]https://www.guru99.com/unsupervised-machine-learning.html

[12]https://www.geeksforgeeks.org/cosine-similarity/

[13]https://builtin.com/data-science/collaborative-filtering-recommender-system

[14]https://towardsdatascience.com/recommendation-systems-a-review-d4592b6caf4b