

# A Machine Learning Approach to Building a Tourism Recommendation System using Sentiment Analysis

Abhishek Kulkarni  
Department of  
Information Technology  
NBN Sinhgad School of  
Engineering, Pune, India

R. M. Samant  
Assistant Professor  
Department of  
Information Technology  
NBN Sinhgad School of  
Engineering, Pune, India

Prathamesh Barve  
Department of  
Information Technology  
NBN Sinhgad School of  
Engineering, Pune, India

Aarushi Phade  
Department of  
Information Technology  
NBN Sinhgad School of  
Engineering, Pune, India

## ABSTRACT

Opinions have become extremely vital in today's "ratings" driven technological services. An android application, a top-tier restaurant or any service for that matter thrives or wanes away on the reviews it gets. A good review can help attract potential users while a bad one may drive them away. Thus, it is essential to analyze these reviews to better understand the user's experience and work towards improving it. The general system that most services use today is based on star-ratings or a score out of 5 or 10. Although these serve the most basic purpose, text-based reviews allow one to understand the reason behind the ratings and are useful to both the user and the service provider to gain more insight. It is impractical for a human to go through thousands of reviews and comprehend the user's sentiment. Instead, training an algorithm to do this job is much more pragmatic and the advances in machine learning allows one to do so. This is where sentiment analysis comes in. In this paper, analysis of various machine learning algorithms like Multinomial Naïve Bayes, Random Forest Classifier and Bernoulli's Naïve Bayes has been done and their behavior has been studied. In addition, study of Convolutional Neural Networks and Recurrent Neural Networks is done to find out if deep learning algorithms perform better. Using these results, a recommendation system is built that maps an individual user's interests to the highest rated tourist places and generates a unique tour plan that is tailored to the user's needs.

## Keywords

Machine Learning, Sentiment Analysis, Tourism, Recommendation System, Recurrent Neural Network

## 1. INTRODUCTION

The tourism industry has exploded in recent years. This explosion has led to the industry becoming more dynamic and user driven. With the advent of new technology, it becomes imperative to entwine the two and create better, more efficient solutions to these dynamic problems. As with any technology, its design revolves around the user. The user's capabilities and proficiency become a major factor which decides the complexity and usefulness of any technology. Various factors like businesses coming online, increase in the quality of

global positioning systems and the popularity of social media has led to the tourism industry becoming multifaceted. In the 20th and early 21st century, planning a trip would take immense efforts on the user's part. This included contacting every hotel for booking information, arranging travel and deciding on various places to visit. The world wide web completely changed the scenario. Tourists came online and wrote reviews of places, businesses came online, and it

became easier to book travel and accommodation. These developments led to ease on the part of the user. Still, the users must manually go through reviews and decide on the best resources among the hundreds available.

The proposed system in this paper eliminates this effort. The proposed system analyses various reviews of tourist places and creates a recommendation list. After taking the interests of the user, the system creates a tailored tour plan for the user. The user can search for information of various tourist places as well as explore resources relevant to her/him.

To generate the said recommendation list, various machine learning and deep learning algorithms have been tried out. These include Bernoulli Naïve Bayes, Multinomial Naïve Bayes, Random Forest Classifier, Recurrent Neural Networks and Convolutional Neural Networks. The basic concept used for this work is Sentiment Analysis. Using existing reviews, the model is trained to identify to what extent a review is positive and negative. This positivity or negativity decides the rating of the tourist place in the system. Further, the recommendation list considers both the rating and user's interests to find the best trade-off between the two.

## 2. LITERATURE SURVEY

In [1] it gives us an outline of intelligent tourism system. Different modules of intelligent tourism system like place recommender system, database thinking, information delivery for tourism etc. are described and their points of interest and breaking points are tended to. It looks at two best recommender framework advancements, Tripplahop's TripMatcher and VacationCoach's Expert Advise Platform, MePrint.

In [2] Hybridization of collaborative filtering and content-based recommendation system is studied. They have utilized IMDB dataset to suggest, having a set of 13 features to recommend a movie. Optimal feature weights are considered, and a regression framework is described. Additionally, execution of the framework is examined.

In [3] multiple content-based recommendation models like TFIDF profile model, BM25 profile model are presented and assessed. So as to investigate the performance of the methodologies these two recommenders are thought about utilizing two diverse datasets acquired from Delicious and Last.fm social frameworks.

In [4] a positioning framework for suggestion of items that gives best incentive to shoppers' cash is proposed to be created. It utilizes novel dataset of US lodging reservation. In view of evaluations from the model, monetary effect of different administration and area qualities of lodging is inferred. In [5] contrasts and subtleties between two distinct methodologies for

text classification, for example Multivariate Bernoulli model and Multinomial model are portrayed. In result it states, Multivariate Bernoulli algorithm performs good on little vocabulary sizes and Multinomial performs better at large vocabulary sizes. Execution of Multinomial Naïve Bayes can be upgraded by utilizing locally weighted learning.[7]

In [6] various ways to deal with make a recommendation list for the travel industry are examined. It expresses that by utilizing content-based scoring, framework can utilize typical tourist media information to include scores-based contents and its semantics to the general inference process.

In [8] Ensemble classification is used to analyze sentiment analysis on twitter dataset. Ensemble classification includes joining the impact of different autonomous classifiers on a specific issue which beats traditional Machine Learning classifiers by 3-5%

[9] This paper actualizes Sentiment Classification task on Amazon Fine Food reviews dataset and Yelp challenge dataset. James Berry thought about two methodologies, first - conventional Bag of Words approach using Multinomial Naïve Bayes and Support Vector Machine Classifiers and second – Long Short-Term Memory (LSTM) Recurrent Neural Network with GloVe Embeddings and self-learned Word2Vec embeddings. This paper concludes LSTM is best performing algorithm.

### 3. PROPOSED METHODOLOGY

#### 3.1 Overview of the System

The proposed system aims to reduce the effort on the user's part. The system will create a recommendation list which is curated using the results of analysis of numerous user-reviews and the inputs given by the user. The deep learning algorithm will determine the extent to which the review of a place is positive and negative. Based on the result, the rank of the place in the recommendation list will be decided. A more positive result will rank the place higher and increase the chances of it being recommended while a more negative result will rank the place lower, thereby decreasing the likelihood of it being recommended. Each of these places is categorized based on what it offers, for instance the Taj Mahal being a

historical site offers a historical and heritage value. A user will enter their preference. This includes the type of location they want to visit (adventurous, historical, architectural, etc.), the number of people traveling and children (if any), and the number of days they plan to take the trip for. Based on these parameters and the user reviews for each place, the recommendation list will be generated uniquely for that user. It will be mapped to the individual user's requirements and a tailored trip will be generated. Thus, the user won't have to settle for generalized plans that tour businesses generally offer.

This system works in two phases. In the first phase the reviews and other relevant data is gathered, and an average rating is assigned to each place. In the second phase, the ratings assigned in the previous phase and other parameters taken from the user are utilized to generate a unique recommendation list. Thus, every user gets a tailored tour plan that actually considers their opinions.

#### 3.2 Design of the Recommendation List

The crux of the system is the recommendation list which maps user's interests to ratings analyzed from reviews. Ratings, ambience, cleanliness, must-visit, nightlife, parking and peacefulness are the factors considered while generating the recommendation list. For features which a user would

generally want, a full score is given. Following formula has been devised for the same,

$$\begin{aligned} \text{Score}(\text{place}=X) = & 10*\text{ambience} + 10*\text{cleanliness} + 10*\text{must-visit} \\ & + \text{nightlife}*\text{NightlifeUserValue} + \\ & \text{parking}*\text{ParkingUserValue} + \\ & \text{peacefulness}*\text{PeacefulnessUserValue} + \\ & \text{childSafety}*\text{ChildSafetyUserValue} + 10*\text{ratings} \end{aligned}$$

Here, the maximum value of each feature is 5, thus the maximum total score of any place will be 400. The features ambience, cleanliness, must-visit status and ratings are preferred by most of the users, so it is assumed that its value will be maximum. For the other features, user's inputs will determine the value. Thus, the summation of these values will result in a holistic score of each tourist place and arranging these scores in descending order will generate the recommendation list.

#### 3.3 Getting Data

Deep learning algorithms used for Sentiment analysis require a vast dataset. For this reason, Amazon Product Reviews dataset from Kaggle having 3.6 million reviews has been used. From this dataset, 1 million reviews are taken. These 1 million reviews contain 600,131 positive reviews and remaining 399,869 negative reviews. Along with these surveys have been gathered of better places utilizing Google API. When one looks through a spot-on Google, Google API returns data about that place alongside 5 most recent surveys for each spot. Additionally, reviews on destinations like TripAdvisor, Google which are openly accessible and are taken to assemble the dataset. Likewise, to get the reviews a survey was conducted getting reviews about various places. Utilizing these, a sum of 30,000 surveys of better places were accumulated. These accumulated surveys are given classification as positive or negative manually. From above dataset, 1 million surveys of Amazon Product Reviews dataset and 20,000 reviews of places are used as training set for algorithm, while remaining 10,000 reviews are used for testing. As a model for the recommendation system, the state of Goa from India is considered. 26 places from Goa are chosen. Values for features like ambience, cleanliness, peacefulness of each spot are given physically by perusing surveys.

#### 3.4 Model Building

To find out the best performing models, the following machine learning and deep learning algorithms were considered and implemented on the Amazon Reviews dataset:

##### I) Bernoulli Naïve-Bayes

In the multivariate Bernoulli event model, features are independent Booleans (binary variables) describing inputs. Like the multinomial model, this model is popular for document classification tasks, where binary term occurrence features are used rather than term frequencies. If  $x_i$  is a Boolean expressing the occurrence or absence of the  $i^{\text{th}}$  term from the vocabulary, then the likelihood of a document given a class  $C_k$  is given by

$$(x|C_k) = \prod_{i=1}^n p_{ki}^{x_i} (1 - p_{ki})^{(1-x_i)}$$

where  $p_{ki}$  is the probability of class  $C_k$  generating the term  $x_i$ . This event model is especially popular for classifying short texts. It has the benefit of explicitly modelling the absence of terms. When implemented on the Amazon dataset, it had an accuracy of 82.75% and an f1-score of 0.83.

##### II) Multinomial Naïve-Bayes

With a multinomial event model, samples (feature vectors) represent the frequencies with which certain events have been

generated by a multinomial ( $p_1, \dots, p_n$ ) is the probability that event  $i$  occurs (or  $K$  such multinomials in the multiclass case). A feature vector  $x = (x_1, \dots, x_n)$  is then a histogram, with  $x_i$  counting the number of times event  $i$  was observed in a particular instance. This is the event model typically used for document classification, with events representing the occurrence of a word in a single document. The likelihood of observing a histogram  $x$  is given by

$$p(x|C_k) = \frac{(\sum_i x_i)!}{\prod_i x_i!} \prod_i p_{ki}^{x_i}$$

If a given class and feature value never occur together in the training data, then the frequency-based probability estimate will be zero. This is problematic because it will wipe out all information in the other probabilities when they are multiplied. Therefore, it is often desirable to incorporate a small-sample correction, called pseudo count, in all probability estimates such that no probability is ever set to be exactly zero. This way of regularizing Naïve Bayes is called Laplace smoothing. Implementing this algorithm on the

Amazon dataset yields an accuracy of 84.48% and an f1-score of 0.85.

### III) Random Forest Classifier

Random Forest learning is the construction of a decision tree from class-labelled training tuples. A random forest is a flow-chart-like structure, where each internal (non-leaf) node denotes a test on an attribute, each branch represents the outcome of a test, and each leaf (or terminal) node holds a class label. The topmost node in a tree is the root node. Classification and Regression Tree (CART), Iterative Dichotomiser 3 (ID3) and Chi-squared Automatic Interaction Detector (CHAID) are few types of decision tree learning algorithms.

The Amazon Reviews dataset when used to train this algorithm outputs an accuracy of 84.60% and an f1-score of 0.85.

A problem arises when using traditional RNNs for NLP tasks because the gradients from the objective function can vanish or explode after a few iterations of multiplying the weights of the network. For such reasons, simple RNNs have rarely been used for NLP tasks such as text classification [7]. In such a scenario one can turn to another model in the RNN family such as the LSTM model. LSTMs are better suited to this task due to the presence of input gates, forget gates, and output gates, which control the flow of information through the network.

**Table 1. Results**

Algorithm Used	Accuracy
Bernoulli Naïve-Bayes	82.75%
Multinomial Naïve-Bayes	84.48%
Random Forest	84.60%
Convolutional Neural Network	94.40%
Recurrent Neural Network	94.56%

Thus, from the analysis it is observed that Recurrent Neural

### IV) Convolutional Neural Network

A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolutional layers, RELU layer i.e. activation function, pooling layers, fully connected layers and normalization layers.

Description of the process as a convolution in neural networks is by convention. Mathematically it is a cross-correlation rather than a convolution. This only has significance for the indices in the matrix, and thus which weights are placed at which index. Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

As expected, the CNN model yielded an accuracy of 94.40%.

### V) Long Short-Term Memory RNN

For the neural network approach, LSTM RNNs have been used because they generally have a superior performance than traditional RNNs. A problem arises when using traditional RNNs for NLP tasks because the gradients from the objective function can vanish or explode after a few iterations of multiplying the weights of the network. For such reasons, simple RNNs have rarely been used for NLP tasks such as text classification. In such a scenario, one can turn to another model in the RNN family such as the LSTM model. LSTMs are better suited to this task due to the presence of input gates, forget gates, and output gates, which control the flow of information through the network.

An accuracy of 94.56% was obtained using this algorithm on the Amazon Reviews dataset.

## 4. RESULTS

For the neural network approach, LSTM RNNs is used because they generally have a superior performance than traditional RNNs for learning relationships.

Network performs the best.

## 5. CONCLUSION

Thus, to develop the recommendation list, various machine learning and deep learning algorithms have been discussed to analyze the reviews of the Amazon Reviews dataset. As can be seen from the evidence above, the Recurrent Neural Network proves to be the model which yields the highest accuracy of 94.56%. Thus, in this experiment a deep learning algorithm outperforms the machine learning algorithms and is consequently chosen to classify the user reviews. The proposed system will thus take the output of this analysis and map it with the user's interests.

In the proposed system, the reviews are looked at holistically. Breaking this review down based on multiple core properties may result in a more in-depth and accurate classification. For instance, in a review about a tourist spot, extracting features like parking availability, cleanliness, child-safety may prove to be helpful and needs further exploration in the future.

## 6. ACKNOWLEDGEMENTS

We greatly acknowledge Amazon Co. and Kaggle for making the dataset of the Amazon Product reviews openly available.

## 7. REFERENCES

- [1] Staab, S., Werthner, H., Ricci, F., Zipf, A., Gretzel, U., Fesenmaier, D. R., ... & Knoblock, C. (2002). Intelligent systems for tourism. *IEEE Intelligent Systems*, (6), 53-64
- [2] Debnath, S., Ganguly, N., & Mitra, P. (2008, April). Feature weighting in content-based recommendation system using social network analysis. In *Proceedings of the 17th international conference on World Wide Web* (pp. 1041- 1042). ACM.
- [3] Cantador, I., Bellogín, A., & Vallet, D. (2010, September). Content-based recommendation in social tagging systems. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 237-240). ACM.
- [4] Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, 31(3), 493-520.
- [5] McCallum, A., & Nigam, K. (1998, July). A comparison of event models for naive bayes text classification. In *AAAI- 98 workshop on learning for text categorization* (Vol. 752, No. 1, pp. 41-48).
- [6] Berka, T., & Plöbñig, M. (2011). Designing recommender systems for tourism. *Proceedings of ENTER 2011*, 26-28.
- [7] Kibriya, A. M., Frank, E., Pfahringer, B., & Holmes, G. (2004, December). Multinomial naïve bayes for text categorization revisited. In *Australasian Joint Conference on Artificial Intelligence* (pp. 488-499). Springer, Berlin, Heidelberg.
- [8] Kanakaraj, M., & Guddeti, R. M. R. (2015, February). Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques. In *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)* (pp. 169-170). IEEE.
- [9] Barry, J. (2017). Sentiment Analysis of Online Reviews Using Bag-of-Words and LSTM Approaches. In *AICS* (pp. 272-274).