Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 12, No.2, December : 2024

COMPARATIVE ANALYSIS OF MACHINE LEARNING APPROACHES FOR SENTIMENT EVALUATION OF HOTEL FEEDBACK

B. N. DHANASEKARAN Assistant Professor Department of Computer Science J.K.K.Nataraja college of Arts & Science, Namakkal, Tamil Nadu bndsekaran@gmail.com,
S. BALAJI Assistant Professor Department of Computer Science J.K.K.Nataraja college of Arts & Science, Namakkal, Tamil Nadu balajisekar2308@gmail.com

Abstract
It is generally recognized that the substantial advancement of Internet technology has significantly transformed the tourism sector, particularly in the realm of hotel reservations. Numerous studies have indicated that a majority of travelers base their choices on online reviews of hotels. Aiming to accurately and effectively forecast hotel guests' satisfaction, this research gathers ten thousand hotel reviews from across the United States, rated between one and five stars, sourced from TripAdvisor. The efficacy of several machine learning algorithms, including Logistic Regression, Naive Bayesian, Decision Tree, Random Forest, Support Vector Machine (SVM), and Neural Network, is evaluated through two experiments utilizing 10-fold cross-validation. Findings reveal that models trained specifically on review titles, which average only 25.7 bytes in length, can attain impressive classification accuracy levels ranging from 84% to 87%. When incorporating review content and filtering out irrelevant words, the SVM model is capable of achieving a peak predictive accuracy of nearly 92%. Additionally, this study illustrates that traditional machine learning algorithms, such as SVM and Naive Bayesian, demonstrate computational efficiency while maintaining strong performance in accuracy, recall, and precision, whereas neural networks necessitate meticulous design of their structure and parameter adjustments.

Introduction

With the swift advancement of Internet technology, the tourism sector, particularly in the realm of hotel reservations, has undergone significant transformation [1]. Data indicates that roughly 150 million tourist reservations were completed online in the year 2016 [2], and this trend has become increasingly intense to date. Additionally, researchers have discovered that ratings from social media reviews can more accurately explain hotel performance metrics compared to conventional customer satisfaction assessments [3]. Furthermore, through the methodology of surveys, a study [4] has disclosed that 70 percent of potential clients globally rely on hotel reviews from websites for their decision-making processes, especially in scenarios where travelers have not visited the destinations before [5]. Hence, extracting textual insights and analyzing sentiments from hotel reviews is advantageous for the entire tourism industry framework.

Outstanding research has been conducted on the textual analytics of hotel guest experiences derived from online reviews [6][7]. Nevertheless, the prevailing tourist experience evaluation systems face several issues: (1) customer feedback can often be informal and ambiguous, meaning that traditional factor analysis and linear models struggle to encapsulate most patterns present in high-dimensional data. (2) With the rise of deep neural networks and the field of Natural Language Processing becoming remarkably popular over the last decade, neural network models containing millions, or even billions, of parameters can deliver exceptional performance [8]. However, the training phase may extend to several months, and there is no assurance that such models will not suffer from overfitting.

In this research, ten thousand hotel reviews were gathered from TripAdvisor, with reviews scoring between 1 to 3 stars classified as ‘negative’ and those scoring 4 to 5 stars identified as ‘positive’. To assess the effectiveness of both classical and contemporary machine learning models, a range of machine learning algorithms, including Logistic Regression, Naive Bayes, Decision Trees, Random Forest, Support Vector Machine (SVM), and Neural Networks, were evaluated through two 10-fold cross-validation experiments to mitigate the likelihood of overfitting. This study seeks to contrast the predictive capability and training efficiency of commonly used machine learning models and explore how text processing influences the performance of these models.

Data Preparation

Collection of Data:

The information was gathered from the database: data.world [9], which obtained and arranged 10,000 hotel reviews from TripAdvisor into csv formats. As shown in Fig.1, nearly 7800 evaluations are rated between 4 to 5 stars and were classified as 'positive (1)', whereas about 2200 evaluations are scored between 1-3 stars and were categorized as 'negative (0)'.

Developmental Methodology

To address the issue of overfitting, the dataset is randomly shuffled and separated into ten segments. In every iteration, 9 segments (9,000 data points) were employed for training, while the labels from the remaining segment (1,000 data points) were used for testing purposes. This methodology is referred to as 10-fold cross-validation, and it aids in mitigating the likelihood of excessive fitting and locally optimal solutions.

As presented in Table 1, since only titles were utilized in the initial experiment, each individual data vector consists solely of a limited number of words, resulting in an average byte size of 25.7 and a standard deviation of 14.6. In the subsequent experiment, the content of the reviews was initially cleansed of punctuation marks (such as commas, periods, semi-colons, colons, hyphens, etc.) and neutral emotive terms that appeared frequently ('hotel', 'inn', 'motel', etc.). Following the incorporation of the filtered review content for training, the mean byte size of each sample rose from 25.7 to 442.4, with a more pronounced variation in sample size, as indicated by a standard deviation of 394.0.

Designs	**Details**		
	Training/Testingdataset size	***Mean ofbytes***	***Standard deviationof bytes***
Experiment 1	9000/1000	26.7	15.6
Experiment 2	9000/1000	452.4	414.0

TABLE 1: PREPRATION OF EXPERMIENTS

Analysis

Analysis of methodology 1:

As illustrated in Table 2, the six machine learning frameworks are sourced from the Python module 'sklearn' [10], and the parameters utilized for the models are detailed comprehensively for the purpose of facilitating experimental replication by readers. For example, in the case of Neural Network, a structure featuring three layers was employed, with the neuron distribution being 5, 4, and 2 across the respective layers. The relu function was selected as the activation function, while the learning rate was set to 10-5.

TABLE II. 'SKLEARN' MODULE PARAMETERS

Models	**Parameters**
Decision Tree	criterion='gini', min_samples_leaf=1,min_samples_split=2

Naive Bayesian	alpha=1.0, fit_prior=True
Logistic Regression	max_iter=100, multi_class='warn',penalty='l2', solver='lbfgs'
SVM	degree=3, gamma='scale', kernel='rbf'
Neural Network	activation='relu', alpha=10-5,hidden_layer_sizes=(5, 4, 2)
Random Forest	criterion='gini', n_estimators=100

The performance metrics for the six models, specifically Decision Tree, Naive Bayesian, Logistic Regression, SVM, Neural Network, and Random Forest, are presented in Fig.2 using a boxplot format, which includes five horizontal lines to represent the minimum, first quartile, median, third quartile, and maximum values within the data. Initially, it is noted that the Neural Network exhibits the poorest performance when compared to the other models, a result attributed to the limited training data, which averages merely 25.7 bytes. Additionally, considering that Random Forest is composed of averaged outputs from 100 decision trees subjected to specific randomness, it surpasses the Decision Tree in accuracy. Thirdly, as indicated previously, the training data is minimal and lacks variety, leading to the observation that linear models such as Logistic Regression tend to outperform non-linear models like Neural Network. Furthermore, SVM advantages itself by effectively handling unstructured data and scaling well to high-dimensional datasets, which gives it the best predictive performance for textual analysis.

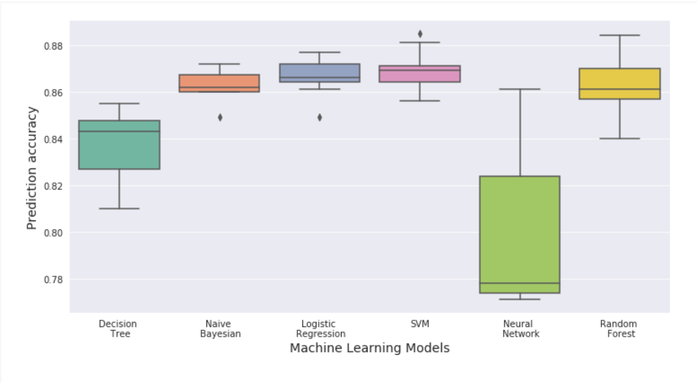

Figure 2. Accuracy performance of ML models on title texts

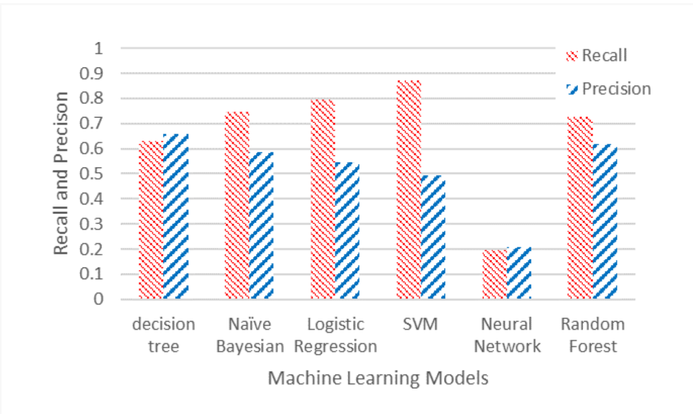

Figure 3. Recall and Precision of ML models on title texts

Last yet importantly, the titles of each review consist merely of a few words, and processing them through Machine Learning models requires only a time frame in seconds. Despite this, five out of the six models manage to attain prediction accuracies ranging from 84% to 87%. This finding implies that even in the absence of datasets on the scale of gigabytes or deep neural networks trained for extensive durations with billions of parameters, traditional models with relatively small structures possess the capability to deliver satisfactory predictive performance. In reference to Fig.3, even though SVM achieves the highest predictive accuracy, this model demonstrates a notable imbalance between precision and recall. Moreover, concerning Logistic Regression, which attains the second-best accuracy, there remains a significant divergence between precision and recall. Consequently, it can be inferred that while both SVM and Logistic Regression models yield commendable accuracy based on limited training sample sizes, there exists a concerning tendency for these models to misclassify positive customer reviews as negative. In practical applications, such tendencies could lead to unintended bias.

Analysis of methodology 2:

As outlined in Section 2.2, a significantly greater amount of textural data was added for training purposes, and disruptive elements such as punctuation marks and commonly used neutral words were eliminated prior to conducting the experiment.

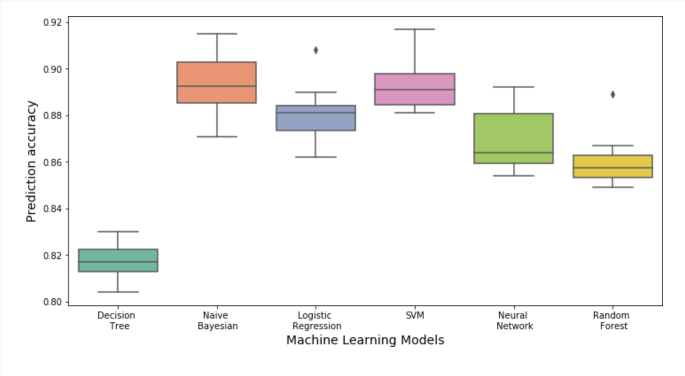

Figure 4. Accuracy performance of ML models on filtered review contents

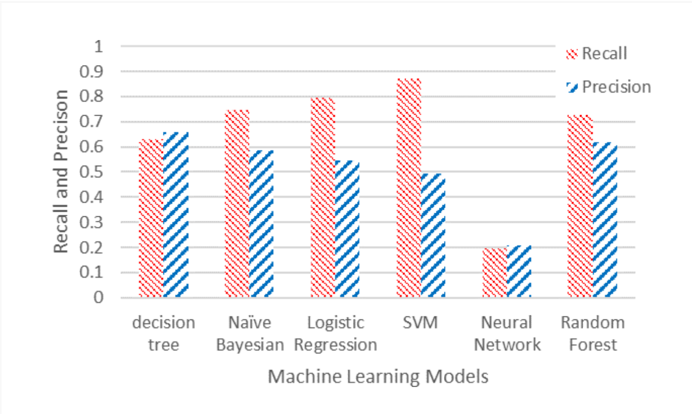

Figure 5. Recall and Precision of ML models on filtered review contents

Based on the variances observed between Fig.4 and Fig.5, it is evident that both SVM and Naive Bayesian models attain nearly 92% peak prediction accuracy alongside approximately 89% median predictive accuracy. Therefore, by supplying a greater volume of pertinent data to machine learning models, there will be a noticeable enhancement in performance. Furthermore, the neural network achieved an accuracy of 86%, which represents an increase of 8% from 78% when only titles were used for training. This observation aligns with the characteristic of Neural Networks that demonstrates improved predictive capability with larger, relevant datasets.

In comparison to Fig.3, Fig.5 reveals that both Logistic Regression and SVM exhibit well-mannered distinctions between recall and precision. Consequently, these models are less prone to bias towards True positives and commit fewer errors in incorrectly categorizing actual positive reviews as negative.

Conclusion:

In this study, a collection of ten thousand English-language hotel evaluations was assembled and classified as either negative or positive. Subsequently, both the titles of the reviews and the curated content of those reviews were utilized to train a variety of models — including Logistic Regression,

Naive Bayes, Decision Trees, Random Forests, SVM, and Neural Networks — through a 10-fold cross-validation process. The findings are encapsulated in the following summary:
When the training samples possess small byte sizes, the 3-layer Neural Network model notably exhibits inferiority compared to the other five traditional models. As the amount of textual data increases, the predictive accuracy of the Neural Network model greatly enhances. Nonetheless, the Neural Network still fails to outperform conventional models like SVM and Logistic Regression across the two tests, which suggests that a careful design of the network structure and parameter adjustments are essential for optimal performance. Out of all six models tested, SVM stands out as the most accurate, precise, and reliable due to its capability to handle unstructured data and high-dimensional attributes effectively. When the models were trained directly on review titles averaging only 25.7 bytes in length, five out of the six models managed to achieve a prediction accuracy between 84% and 87%. Given that the training process requires only a matter of seconds, this result highlights the potential for applying traditional machine learning models within individual hotels to gauge customer satisfaction. With an increase in the volume of textual data, alongside the removal of irrelevant punctuation and neutral words, the SVM model could attain a peak predictive accuracy of nearly 92%.

REFERENCES

[1] A. Emir, H. Halim, A. Hedre, D. Abdullah, A. Azmi, and S. Kamal. Factors influencing online hotel booking intention: A conceptual framework from stimulus-organism-response perspective. International Academic Research Journal of Business and Technology, 2016, 2(2), 129-134.
[2] Statistic Brain, 2016. Internet Travel & Hotel Booking Statistics. https://www.statisticbrain.com/internet-travel-hotel-booking-statistics/
[3] W. G. Kim, and S. A. Park. Social media review rating versus traditional customer satisfaction: which one has more incremental predictive power in explaining hotel performance?. International Journal of Contemporary Hospitality Management, 2017, 29(2), 784-802.
[4] A. J. Flanagin, and M. J. Metzger. Trusting expert-versus user-generated ratings online: The role of information volume, valence, and consumer characteristics. Computers in Human Behavior, 2013, 29(4), 1626-1634.
[5] R. K. Nielsen, and K. C. Schrøder. The relative importance of soc ial media for accessing, finding, and engaging with news: An eight-country cross-media comparison. Digital journalism, 2014, 2(4), 472-489.
[6] X. Tian, W. He, R. Tao, and V. Akula. Mining online hotel reviews: a case study from hotels in China, 2016.
[7] Z. Xiang, Z. Schwartz, Jr, J. H. Gerdes, and M. Uysal. What can big data and text analytics tell us about hotel guest experience and satisfaction?. International Journal of Hospitality Management, 2015, 44, 120-130.
[8] Z. Jianqiang, G. Xiaolin, and Z. Xuejun. Deep convolution neural networks for twitter sentiment analysis. IEEE Access, 2018, 6, 23253-23260.
[9] data.world, 2019. Hotel Reviews – dataset in DATAFINITY. https://data.world/datafiniti/hotel-reviews.
[10] J. Hao, and T. K. Ho. Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. Journal of Educational and Behavioral Statistics, 2019, 44(3), 348- 361.