

RESEARCH ARTICLE



OPEN ACCESS

Received: 06.05.2021

Accepted: 20.10.2021

Published: 02.12.2021

Citation: Chigateri K, Bhandarkar R (2021) Necessity and Preference Mining from Text Reviews: A Non-bipolar assessment of Text Reviews. Indian Journal of Science and Technology 14(40): 3064-3073. <https://doi.org/10.17485/IJST/v14i40.701>

* **Corresponding author.**

keerthanabc@nitte.edu.in

Funding: None

Competing Interests: None

Copyright: © 2021 Chigateri & Bhandarkar. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](https://www.indst.org/))

ISSN

Print: 0974-6846

Electronic: 0974-5645

Necessity and Preference Mining from Text Reviews: A Non-bipolar assessment of Text Reviews

Keerthana Chigateri^{1*}, Rekha Bhandarkar²

¹ Assistant Professor, CSE Department, N.M.A.M Institute of Technology, NITTE, Udupi, Karnataka, India

² Professor and Head, ECE Department, N.M.A.M Institute of Technology, NITTE, Udupi, Karnataka, India

Abstract

Objective: To perform opinion mining on text reviews related to hotel. **Methods:** In this work, the opinion is mined by identifying and extracting necessities and preferences along with the associated two features or aspects expressed in text reviews by customers. The hotel dataset (From Kaggle website, hotels in United States, has 35912 samples) is considered for training and testing. Modals 'Has' and 'Would' are used to identify and extract reviews which are expressing the necessities and preferences of customers from the dataset of hotel reviews. Random Forest machine learning algorithm method is used for classifying the reviews belonging to necessity and preference categories. **Findings:** From the related works carried out so far, it is indeed transparent that so far, the text reviews are analysed for general sentiments like good, bad etc., polarities like positive, negative or neutral and emotions like joy, fear etc., The analysis for necessities and preferences in the text is yet to be addressed. The current research focuses on narrowing the semantic gap in opinion mining from Generalized analysis of reviews like positive, negative, good, bad to Specialized analysis of reviews like mining necessities and preferences of customers which may give higher level of understanding of customer needs by service providers. In this work, the reviews are classified into two classes viz, necessities and preferences are identified and classified using Random Forest machine learning algorithm. It gave the accuracy of 91% in classifying the reviews as necessity and 99.78% in classifying the reviews as preferences by using the formula given in the system implementation section. **Novelty:** Classification of reviews into Necessity and preference classes. **Keywords:** Reviews; Opinion mining; Necessities; Preferences; Modals; Random Forest

1 Introduction

In this work filtering and categorising of hotel reviews based on modals are carried out and this is further used for opinion mining. Here to deal with reviews of hotel, basically hotel reviews dataset is considered. This is pre-processed to get only those reviews which

has modals in it and in order to get that, at first all the punctuations, numbers, unwanted words are removed. Only reviews which have Modals in it are considered. Further these reviews are categorized into two different classes based on different modals. Reviews with each type of required modal is considered in particular files and later the model is trained by giving these review files with different modals as input. The proposed method for feature-based opinion mining for Hotel reviews has several steps. The various steps involved in this opinion mining are input reviews collection, input data pre-processing, feature extraction along with the opinion indicating modal extraction, and result summarization which are described below. The proposed model for the effective feature-based opinion mining is depicted in the figure below. The input data contains reviews about the features of the Hotel i.e., about room, food, WIFI Service etc. It has mainly five phases. Each phase depends on the output of other phase and it proceeds as a series. Pre-processing refers to the removal of unwanted data from the input data sets. The first step in pre-processing is to split the extracted reviews into sentences and then to a list of words. This process is called tokenization. Further, stop words like “a”, “an”, “the” which are not required for the opinion classification are removed from the tokenized data. Finally, tokenized data are tagged using Parts of Speech tagger (POS).

The algorithm used in this work is Random forest. Like its name implies, it consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes become our model's prediction.

In this work non-bipolar assessment of text reviews is carried out. Bi means two. Two Polarities are positive and negative. Bi-polar assessment of reviews means classifying the reviews universally as positive or negative. In Tri-polar, tri means tree. Polarities are positive, negative and neutral. In Tri-polar assessment of reviews, reviews are classified universally as positive, negative or neutral. But an entire review may be a combination of positive, negative or neutral sentiments. Some features in the review maybe positive, some maybe negative and/or some maybe neutral. So under such situations labelling entire review with only one polarity may not be appropriate. So non-bipolar assessment of reviews is classifying the reviews beyond these three polarities. Here reviews are classified at feature level as necessity and preference which may be more informative and maybe helpful for different stakeholders. The following sections contain the elaborated explanation related to this.

A novel approach to extract features and opinion words referring to features from the text reviews was presented. These opinion words were classified as positive or negative which is bi-polar classification⁽¹⁾. A general process for categorizing the product reviews collected from Amazon.com as positive, negative or neutral at sentence-level and review-level was proposed. It has used scikit-learn open-source machine learning package in python. Here tri-polar classification of text reviews is carried out⁽²⁾. A novel multi-text summarization technique was proposed for identifying the top-k most informative sentences of hotel reviews Here hotel reviews were classified as positive or negative by machine learning method. This classification and summarization were to help web users to understand review contents easily in short time. This was also bi-polar classification of text reviews⁽³⁾. Customer reaction on to two types of posts (photos or videos) of CocaCola and PepsiCo users on six social networks: Facebook, Twitter, Instagram, Pinterest, Google and YouTube were aimed to analyse. The output of this study was the conceptualization and measurement of a brand's SM (Social Media) ability to understand customer preferences for different types of posts by using various statistical tools and the sentiment analysis (SA) technique applied to big sets of data. This work is related to study of differences in user's sentiments and emotions expressed via photos and videos in different social medias but not text reviews⁽⁴⁾. A method to identify and cluster customers' true needs from customer complaints by using the concept of job in the Outcome-Driven Innovation (ODI) method was suggested. It provided a method to analyse customer complaints by using the concept of job. The ODI-based analysis contributes to identification of customer latent needs during the pre-execution and post-execution steps of product use by customers that previous methods couldn't discover. So this work is identifying customer needs by analysing customer complaints⁽⁵⁾. The biometric information fusion approach as a complement of traditional Web Usage Mining (WUM) techniques to understand the navigational behaviour and preferences of Web users was presented in. This work is a survey of how biometric information fusion is applied to the mining of web usage by extracting the knowledge of web users' interests, navigational actions and preferences from web and biometric data⁽⁶⁾. A recommender system based on user preferences and constraints was proposed. The AprioriTid algorithm and association rules are used to find frequent item sets and preferences to recommend products to users. This work is an extension of online application developed previously for online shopping⁽⁷⁾. The proposed work is for both mining and classification of hotel reviews as preference and necessity using modals and thus its differs from works in⁽¹⁻⁷⁾. A novel approach for mining preferences from user log data based on the concept of strict partial order preferences was presented. So here preference mining is carried out on log data to help personalized e-applications to gain knowledge on customer's preferences for qualified customer service⁽⁸⁾. A preference-based interactive document summarisation framework was proposed which interactively learns to generate improved summaries based on user preferences using active learning to query the user, preference learning to learn a summary ranking function from the preferences, and neural Reinforcement learning to efficiently search for the (near-)optimal summary. Here summary is generated related to a particular document/topic based on user interest by takin user's feedback as input interactively. This

work is at document level and is a summarization tool⁽⁹⁾. The proposed work is for mining both preference and necessity from hotel reviews using modals method. Unlike summarization, it identifies the features which are of necessity and preferred by users and thus it differs from⁽⁹⁾ in dataset and methodology. Evidentially Reasoning(ER) to analyse sentiment words in social networks to investigate consumer preferences was proposed. Here criteria for prioritization (usually by expert's/ key opinion leaders) is identified. Sentiment analysis is carried out by tallying the number of occurrences of sentiment words in the comments. The polarity and sentiment values are determined using SentiWords. It's an approach to summarize user generated content from social networking media⁽¹⁰⁾. The proposed work is for mining both preference and necessity from hotel reviews using modals method and thus it differs from⁽¹⁰⁾. The overall sentiment rating prediction for a review has been shown to improve by capturing facet level rating of author preferences in rating prediction. In this work, the objective is to predict the overall rating of the review as a function of the facet specific opinions weighed by the author's facet-specific preference⁽¹¹⁾. So⁽¹¹⁾ is related to rating prediction and the proposed work is about mining necessities and preferences using modals. The designing and developing of an application to extract opinions from reviews and generate summarization charts was proposed. It's an extension of Liu's aspect-based opinion mining methodology in order to apply it to the tourism domain. Here sentiment orientation for each aspect is determined by rule-based approach offered by Ding, Liu and Yu. Finally, opinions about aspects are summarized as positive or negative using bar charts⁽¹²⁾. So in⁽¹²⁾ bi-polar classification of aspects is done and in proposed work is mining necessities and preferences using modals. A framework which integrates a collaborative filtering approach and an opinion mining technique for movie recommendation was proposed. Sentiment analysis is first applied to the users' reviews to detect consumer opinions about the movie they have watched. sentiment analysis is applied to user reviews to detect consumer's opinions about movies watched by them. Using this analysis preference profile for individuals is built in⁽¹³⁾ and proposed work is mining necessities and preferences using modals. Students prefer university spaces with actual greenery or nature posters, and that they also expect that a green outdoor university environment can be more restorative was showed in⁽¹³⁾. This work is related to benefits students get by interacting and integrating with greenery. Points on consumer buying behaviour and their preferences in the purchase of branded batter through the structured questionnaire to conduct the study was discussed in⁽¹⁴⁾. In⁽¹⁴⁾ how the transformation in marketing strategy needs to be done when a new batter brand enters the segment to develop their business with other competitors and also factors influencing the buyers to buy branded and non-branded batter. In⁽¹⁵⁾, to know the preference of customers for various parameters of selecting an eyewear and to find the perception of customers on significant parameters with reference to branded and local spectacles, a survey by asking questions from people of Nagpur using eyewear for their day-to-day usage was conducted. Questions pertained to factors affecting choice of spectacles and preference level of individuals for branded and local spectacles based on demographic factors was done in⁽¹⁵⁾. So in⁽¹⁴⁾ and⁽¹⁵⁾ preferences are identified in batter and spectacles respectively through questionnaire to users and proposed work is mining necessities and preferences using modals and thus differs from work in^(8,14,15). The factors like significance, consequences, urgency, dominance of opinion mining and its categories, approaches, uses, libraries are examined in^(16,17).

So from the above related work, it is indeed transparent that so far the text reviews are analysed for sentiments like good, bad etc., polarities like positive, negative or neutral and emotions like joy, fear etc., and analysis for necessities and preferences in text is yet to be addressed. Following are some points related to the same:

1. The current research focuses on narrowing the semantic gap in opinion mining from Generalized analysis to Specialized analysis.
Generalized analysis: universally labelling the reviews based on polarity as positive or negative or neutral or classifying the reviews based on emotions like Happy, sad, angry, shocked etc., Here mining of reviews which contain necessities and preferences is yet to be addressed
Specialized analysis: Here mining of reviews which contain necessities and preferences is carried out.
2. Specialized analysis may have higher level of understanding of text reviews than generalized analysis like positive, negative, neutral, happy, sad etc., This kind of analysis may provide information for various stakeholders like customers, business analysts, service providers, marketing, sales, customer-support to decrease the response time, manage limited resources effectively, locate wastage points, trade-off the demand-supply, identify area of concerns, learn future trends.
3. The polarity label or emotion label may be the consequence of necessities and preferences being fulfilled or not fulfilled. For example, if necessities and preferences are met, then the customer may review as positive or happy else negative or sad. So this work is trying to bridge the semantic gap slightly more than the existing polarity and emotion recognitions of reviews by mining necessities and preferences reviews instead of positive/negative/neutral or happy/sad etc.
4. The dataset containing reviews with modal verbs "would prefer/like" and "has" is needed to carry out analysis of reviews carrying necessities and preferences.

Following research questions may arise from related work:

1. It is indeed transparent from related work that so far text reviews are analysed for sentiments like good, bad etc., polarities like positive, negative or neutral and emotions like joy, fear etc., So What other information other than sentiments and emotions the Reviews may contain?
2. Reviews may also contain facts, suggestions, necessities and preferences, comparisons. How analysis of necessities and preferences may be useful for different stakeholders in the society?
3. What methodology may be used to mine necessities and preferences reviews?
4. Is the suitable dataset available to mine necessities and preferences?

In the following ways we would like to propose to respond:

1. reviews may also contain facts, suggestions, necessities and preferences, comparisons.
2. analysis for necessities and preferences in text reviews may be useful for various stakeholders in the society like customers, business analysts, service providers, marketing, sales, customer-support to decrease the response time, manage limited resources effectively, locate wastage points, trade-off the demand-supply, identify areas of concern and novel patterns, learn future trends, to create awareness and give guidance.
3. modal verbs “has” and “would” may be used to mine reviews with necessities and preferences.
4. identifying a dataset of text reviews with sentences containing modal “would” and “has”.

2 System Design

The dataset is collected from Kaggle⁽¹⁸⁾. In this work two aspects viz effective phone answering and room service are considered for preference and necessity categories respectively. Modals along with infinitives and verbs, “has” and “would like to / prefer to / would rather” are used in this work to recognize the reviews carrying necessities and exhibiting preferences respectively from the whole other reviews. Modal is a helping verb. They are used to express different behaviours by adding meaning to the base verb and are used in a sentence before the different forms of verb depending on the context. They are used as it is and are not changed to present, past and future. By and large, modals most common value is futurity when used in the first and third persons. These are no past forms; they refer to the future. Necessity means something which is not optional. ‘Has’ modal is used for expressing the first category necessity informally in present or future. Infinitive is used to indicate something which can be a recurring or one-time need. It’s in the following structure:

Noun/pronoun + Has (modal) + infinitive verb (to) + base verb

The second category considered in this work is preference. Preferences are expressed when different choices are possible. Modal ‘would’ is used for identifying the preference. It is having the following structure:

Noun/pronoun + Would(modal) + base form of the verb + infinitive (to)

The reviews containing these modals are categorized as necessity and preference. Table 1 shows the modals and sentences used for Necessity and Preference categories. This kind of classification may be targeted to service providers as it may help them to provide customized services to customers. Categorization of preference and necessity reviews is followed by feeding only these reviews into two separate files. This separates the less informative reviews from more informative preference and necessity reviews. Processing too general reviews like “it’s good”, “it’s bad” etc., which are not more informative may lead to more time consuming, storage consuming and may not help much in decision making. Instead processing reviews from these files may lead to less storage, processing-time and may provide more information and thus help to make better decisions to serve in customized way as such reviews are expressing preferences and necessities. The service provider can make use of these files for improving their service. All other reviews like “it’s good” are replied as thank you.

Table 1. Modals and Sentences used for Necessity and Preference category

Sl no	Modal	Sentences	Category	Target
1	Would prefer/like	1. We would prefer good lighting for effective work 2. We would like to have AC rooms	Preference	Service provider
2	Has to	1. Cleanliness has to be there 2. Privacy has to there	Necessity	Service provider

3 System Implementation

The block diagram of system implementation is shown in Figure 1. It carries the following steps:

1. Collecting and Pre-Processing the Dataset
2. Categorizing the reviews based on its modals
3. Training and Testing each categories of reviews in Random Forest Algorithm
4. Evaluating the model
5. Predicting the Output
6. Saving and Loading the model



Fig 1. Block diagram of necessity and preference mining

The system implementation consists of six steps which are explained below:

- **Collecting the Dataset:** This is one of the main steps of this analysis. For this work the hotel dataset is taken from the Kaggle. Even though the hotel dataset is used, this model can classify all reviews as necessity and preference based on modals.
- **Pre-Processing the Dataset:** This is the major step after collecting the dataset. It will pass through the following steps explained below. Figure 2 shows the steps involved in pre-processing the dataset:

Pre-processing of data

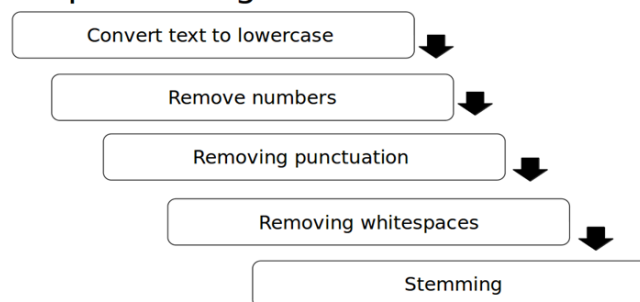


Fig 2. Steps in pre-processing the dataset

- **Removing the numbers from each review:** In this step all the numbers which are present in the reviews are removed as they are not the target for the desired output and it simplifies the review to only text form.
- **Removing the white spaces from each review:** When the user will give input for the review, it may have the empty spaces and which is not required for processing it. It is removed.
- **Removing the punctuations from each review:** Most of the review consists of punctuation marks, which is also not needed for processing, it removed.
- **Converting the review into lower-case and stemming:** All text reviews are converted into same lower case to stabilize the form of review.

Parts of Speech tagging: In this step each review is passed as input and it will give parts of speech with its respective tag. MD is the tag used for modal which are focused in this work.

- **Categorization of reviews based on its modals:** The reviews which are pre-processed are then categorized based on the modals which are obtained from POS tagging. Two categories are made based on modals. First category consists of the modal “has”. Second category consists of the modal “would like to / prefer to / would rather”. And these reviews are then saved into respective files for further processing of data in it.
- **Training and Testing each category of reviews in Random Forest Algorithm:** Before training and testing the dataset, the positive values are specified as Zero (0) in the data set and other all reviews are considered as negative values that is One (1) for the respective categories. It's done to split the dataset into training and testing sets of data. Then these categories are passed as input for Random Forest Algorithm for training and testing the dataset.

Random forest is the powerful supervised machine learning algorithm. It's basically a bagging technique. It creates a forest of number of decision trees. Decision trees are easy to build, use and interpret. They work great with the data used to create them but not flexible for classifying new samples which might lead to inaccuracy. But the Random forest combine the simplicity of decision trees with flexibility resulting in vast improvement in accuracy. In general, more trees in forest more robust the prediction and thus higher accuracy. Here there are multiple trees. For each tree subset of dataset is given as input for training and prediction.

Consider a dataset 'D' with 'd' number of records/rows and 'm' columns. In bagging we have many base learner models say M_1, M_2, \dots, M_n . In random forest, each of this model is a decision tree. Some D' rows (row sampling) and 'n' columns (feature sampling) are picked from 'D' and are given to M_1 as training data such that $D' < D$ and $n < m$. similarly the training data is given to other models also. Some records might appear in more than one model.

Decision trees have low bias and high variance. Low bias is the model is able to training data well. If the decision trees are created to its complete depth, then it gets trained properly for training dataset and thus training error may be very less. High variance is performance is good for training dataset but low for test dataset. For a new test data decision tree might give more error i.e., it may lead to overfitting. So single decision tree may be over fitted but random forest is not. This high variance is converted to low variance in random forest when multiple decision trees are used. This leads to better accuracy for new input. To classify new input based on the attributes each tree gives a classification and it is said tree votes for that class. The forest chooses the classification having the most votes over the other trees in the forests. Thus the decision is aggregated. This method of picking many base learners and aggregating the decision is called as bagging. It can handle the missing values and maintain the accuracy when large proportion of data is missing. When there are more trees in the forest random classifier won't over fit the model. It has the power to handle the large datasets with higher dimensionality

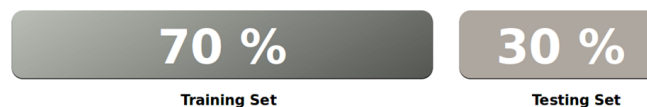


Fig 3. Division of the dataset into training and testing sets

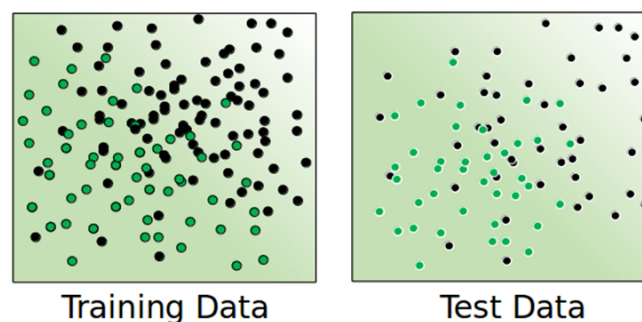


Fig 4. Classification for training and test data

- **Evaluating the Model:** The model is evaluated based on three metrics namely Recall, Precision and F1 Score method. The formulae used for calculating these are as below:

$$\text{Precision } P = TP / (TP + FP)$$

$$\text{Recall } R = TP / (TP + FN)$$

$$\text{F1 score is } F = 2 * ((\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}))$$

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Confusion matrix:

Table 2. Confusion matrix

	Class 0 Predicted	Class 1 Predicted
Class 0 Actual	TP	FN
Class 1 Predicted	FP	TN

Here,

- Class 0: Positive FP= False Positive
- Class 1: Negative FN= False Negative
- TP: True Positive TN= True Negative
- **Predicting the Output:** When the user interacts with our model, the user will be asked to enter the review. The user entered review will be taken as consideration for the processes which are mentioned above. By following all the steps, it will generate the respective review and it will display the category. And the review is then appended in the respective category of file for the further use. For the next iteration the previous review is also a training input.
- **Saving and Loading the model:** In this step all the results and calculations are saved in the desired file and then stored for the future analysis.

4 Results and Discussions:

The following Figures 5, 6, 7 and 8 and Table 3 show the findings of this work such as Preference category output, Necessary category output, evaluation metrics viz accuracies of Preference and Necessity categories and Summarization of results of previous related works respectively.

The accuracy, precision and recall are calculated using the formula mentioned above.

Figures 9 and 10 show the Necessity category output with the aspect/feature Room service and Preference category output with aspect/feature Effective Phone Answer

```
Enter your review: we would rather have veg food
This review belongs to Preference category
Thank you visit again
```

Fig 5. Preference category output

```
Enter your review: Privacy has to be there
This review belongs to Necessity category
Thank you visit again
```

Fig 6. Necessary category output

```
[[ 445 228]
 [ 21 2009]]
```

	precision	recall	f1-score
0	0.95	0.66	0.78
1	0.90	0.99	0.94
accuracy			0.91

Fig 7. Necessary category evaluation metrics

```
[[ 16 13]
 [ 0 2674]]
```

	precision	recall	f1-score
0	1.00	0.55	0.71
1	1.00	1.00	1.00
accuracy			1.00

Fig 8. Preference category evaluation metrics

Enter your review: room service has to be there
This review belongs to Necessity category
Feature is Room Service
Thank you visit again

Fig 9. Necessary category output with aspect/feature Room Service

Enter your review: we would like to have effective phone answering
This review belongs to Preference category
Feature is Effective Phone Answer
Thank you visit again

Fig 10. Preference category output with aspect/feature Effective Phone Answer

Table 3. Summarization of results of previous related works

Sl no.	Previous works	Categories and/or dataset	Performance Metric
1.	Work in this paper	Preference and Necessity Hotel reviews dataset	Accuracy=0.91 and 1.00
2	Feature Specific Sentiment Analysis for Product Reviews	Positive or negative 2 datasets: 500 and 2500 reviews each	80.98
3	Sentiment analysis using product review data	positive, negative or neutral Amazon.com dataset	0.90-0.98
4	Opinion mining from online hotel reviews – A text summarization approach	top-k most informative sentences. online hotel reviews. considered only nouns, adjectives, and negative adverbs for sentiment analysis.	Mean and Standard deviation.
5	Customer Complaints Analysis Using Text Mining and Outcome-Driven Innovation Method for Market-Oriented Product Development	Prepare and confirm Customer complaints	4%
6	Preference Mining: A Novel Approach on Mining User Preferences for Personalized Applications	Log data- categorical, numerical, complex	Categorical: 60 % precision, 38 % recall, Numerical: 58 % precision 40 % recall Complex: 55 % precision 15 % recall.
7	Identifying Customer Preferences about Tourism Products using an Aspect-Based Opinion Mining Approach	TripAdvisor positive, negative	Precision & Recall= 90%
8	A Study on Customer Preference Between Local Spectacles and Branded Eyewear.	Local vs branded spectacles 350 data samples from Nagpur City	Local= higher rating on reasonable price, fashionable, trendy easy availability and fitting. Branded= higher rating on good frames, lenses and comfort level parameters.

5 Conclusion and Future work

In this work, mining of necessities and preferences along with one aspect for each from text reviews is carried out on the hotel dataset obtained from Kaggle. Modals “has” and “would” are used to identify and classify reviews carrying necessities and preferences respectively. It has the accuracy of 91% and 99.78 respectively (Figure 7 and Figure 8). In future, the number of aspects may be increased and extended to other required domains. It may give better information to service providers, decrease the response time and immensely favour traders to fade the barriers between them and customers.

References

- 1) Mukherjee S, Bhattacharyya P. Feature Specific Sentiment Analysis for Product Reviews. *Computational Linguistics and Intelligent Text Processing*. 2012;p. 475–487. Available from: https://doi.org/10.1007/978-3-642-28604-9_39.
- 2) Fang X, Zhan J. Sentiment analysis using product review data. *Journal of Big Data*. 2015;2(1). Available from: <https://dx.doi.org/10.1186/s40537-015-0015-2>.
- 3) Hu YH, Chen YL, Chou HL. Opinion mining from online hotel reviews – A text summarization approach. *Information Processing & Management*. 2017;53(2):436–449. Available from: <https://dx.doi.org/10.1016/j.ipm.2016.12.002>.
- 4) Păvăloaia VD, Teodor EM, Fotache D, et al. Opinion Mining on Social Media Data: Sentiment Analysis of User Preferences. *Sustainability*. 2019;11(16):4459. Available from: <https://dx.doi.org/10.3390/su11164459>.
- 5) Joong J, Jung K, Ko S, Kim K. Customer Complaints Analysis Using Text Mining and Outcome-Driven Innovation Method for Market-Oriented Product Development. *Sustainability*. 2018;11(1):40. Available from: <https://dx.doi.org/10.3390/su11010040>.
- 6) Slanzi G, Pizarro G, Velásquez JD. Biometric information fusion for web user navigation and preferences analysis: An overview. *Information Fusion*. 2017;38:12–21. Available from: <https://dx.doi.org/10.1016/j.inffus.2017.02.006>.
- 7) Mohammed B, Mouhoub M, Alanazi E, Sadaoui S. Data Mining Techniques and Preference Learning in Recommender Systems. *Computer and Information Science*. 2013;6(4). Available from: <https://dx.doi.org/10.5539/cis.v6n4p88>.

- 8) Holland S, Ester M, Kießling W. Preference Mining: A Novel Approach on Mining User Preferences for Personalized Applications. In: Knowledge Discovery in Databases: PKDD 2003. Springer Berlin Heidelberg. 2003;p. 204–216. Available from: https://link.springer.com/chapter/10.1007/978-3-540-39804-2_20.
- 9) Gao Y, Meyer CM, Gurevych I. Preference-based interactive multi-document summarisation. *Information Retrieval Journal*. 2020;23(6):555–585. Available from: <https://dx.doi.org/10.1007/s10791-019-09367-8>.
- 10) Ng CY, Law KMY. Investigating consumer preferences on product designs by analyzing opinions from social networks using evidential reasoning. *Computers & Industrial Engineering*. 2020;139:106180. Available from: <https://dx.doi.org/10.1016/j.cie.2019.106180>.
- 11) Mukherjee S, Basu G, Joshi S. Incorporating author preference in sentiment rating prediction of reviews. In: Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion. ACM Press. 2013. Available from: <https://dl.acm.org/doi/10.1145/2487788.2487805>.
- 12) Marrese-Taylor E, Velásquez JD, Bravo-Marquez F, Matsuo Y. Identifying Customer Preferences about Tourism Products Using an Aspect-based Opinion Mining Approach. *Procedia Computer Science*. 2013;22:182–191. Available from: <https://dx.doi.org/10.1016/j.procs.2013.09.094>.
- 13) van den Bogerd N, Dijkstra SC, Seidell JC, Maas J. Greenery in the university environment: Students' preferences and perceived restoration likelihood. *PLOS ONE*. 2018;13(2):e0192429. Available from: <https://dx.doi.org/10.1371/journal.pone.0192429>.
- 14) Sampathkumar M. Sampathkumar M. Transformation in Marketing – Understanding Consumer Preferences and Buying Behaviour of Branded Idli / Dosa Batter. 2020. Available from: <http://www.jcreview.com/?mno=105286>.
- 15) Polke N. A Study on Customer Preference Between Local Spectacles and Branded Eyewear. *Journal of Critical Reviews*. 2020;7(19). Available from: <http://www.jcreview.com/?mno=135356>.
- 16) What is opinion mining & why is it essential?. Available from: <https://monkeylearn.com/blog/opinion-mining/>.
- 17) Opinion Mining with the Text Analytics Client Libraries. 2019. Available from: <https://devblogs.microsoft.com/azure-sdk/opinion-mining/>.
- 18) Mostipak J. Hotel booking demand. 2019. Available from: <https://www.kaggle.com/jessemostipak/hotel-booking-demand>.