Aspect-based Opinion Mining on restaurant reviews

Group: T2210018
Supervisor: [Farig Yousuf Sadeque], [Moin Mostakim]
Department of Computer Science & Engineering, Brac University

Introduction

- Online review helps customers in terms of decision-making.
- It saves time for the customers to choose restaurants.
- But sometimes high-quality reviews get mixed up with vague ones.
- Opinion mining converts people's thoughts, and emotions toward a restaurant-based sentiment.
- Document-level opinion looks upon allocating total sentiment over the whole review.
- Sentence-level opinion searches the sentiments in each of the sentences in the review.
- An opinion level where the analysis is done on every word in the reviews as the polarity can be different in different situations.

Dataset

- The dataset composed of 2.7 million reviews from 6 different cities
- Having a good amount of reviews and fewer null values makes the dataset ideal for use.
- Some of the columns like parse count, restaurant name, review id, and review preview are removed from the dataset
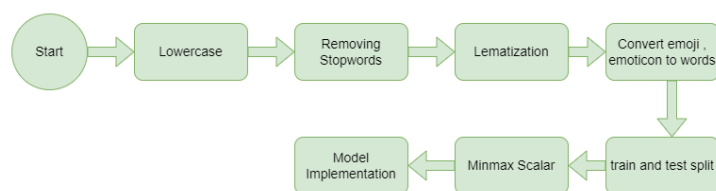

Figure 1. Dataset analysis steps

- After that some of the pre-processing techniques used are lowercase, removing stopwords, expanding contractions, lemmatization, and converting emojis and emoticons into words.
- Finally the dataset is converted into 80% and 20% training and testing.
- And scaling the features using minimax scaler

Literature Review

- Opinion mining doesn't require a deep understanding of text, unlike its counterpart, standard syntactical NLP. It mainly focuses on semantic inference and affective information associated with natural language whereas Syntactical NLP focuses on summarization and auto-categorization [1].
- According to [2], the authors demonstrated the task of aspect-based sentiment analysis by modeling the interdependencies of sentences in a review with a hierarchical bidirectional LSTM.
- The core tasks in aspect-based opinion mining is aspect identification, aspect-based opinion word identification, and its orientation detection [3].
- In this study [4], the researchers classified documents by overall sentiment instead of topics using movie reviews as data and they found that standard machine learning techniques outperform human-produced.

Research objective

- Form an effective model to search and extract all the aspects using NLP techniques.
- Perform sentiment analysis on all the aspects extracted from the reviews.
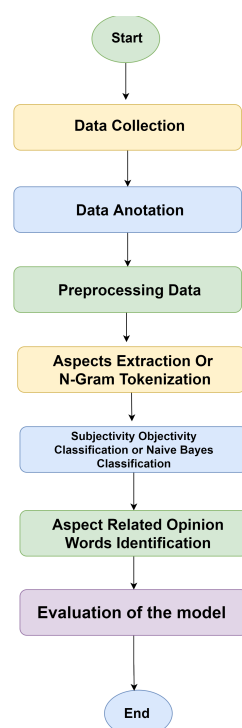- Identify a mapping between the opinions and the extracted aspects.

Figure 2. Another figure caption.

Topic Modelling

Topic modelling is an unsupervised learning algorithm and is a technique thath finds and assigns topics from a given corpus of present words.
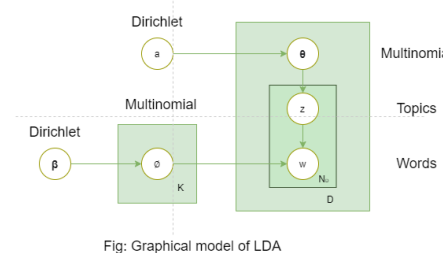
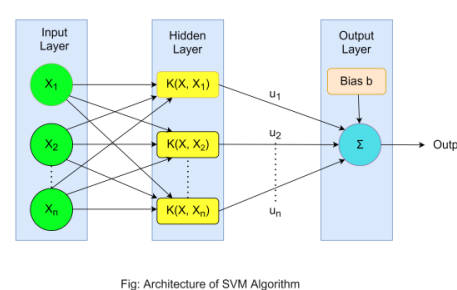

Figure 3. Architechture of LDA.

Models

Figure 4. Architechture of SVM.

- SVM is a linear model for regression and classification that works well for numerous contemporary concerns and is effective for both linear and non-linear problems.
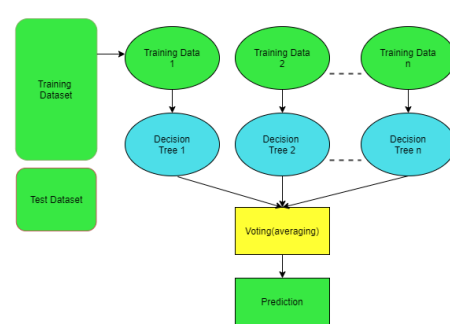

Figure 5. Architechture of random forest algorithm.

- The resultant data of multiple decision trees are merged by the Random Forest to produce a single outcome.
- The MNB algorithm is a supervised probabilistic learning method that is primarily employed in NLP for lexical classification.
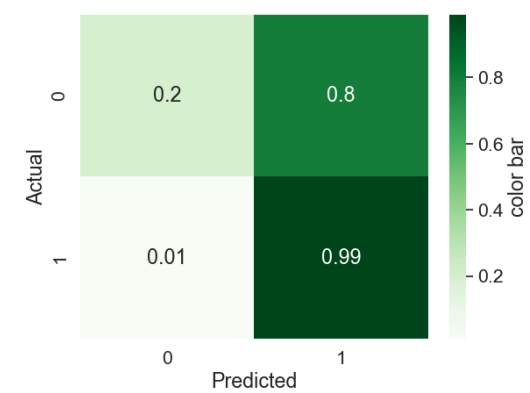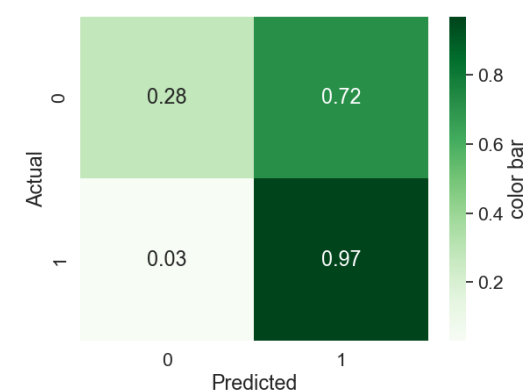
Graph

Figure 6. SVM confusion matrix.

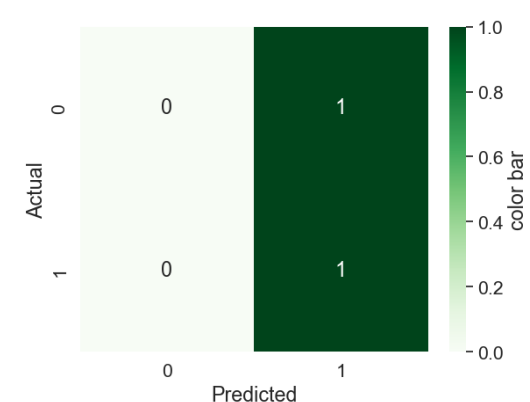Figure 7. Random Forest confusion matrix.

Figure 8. Multinomial naive bayes confusion matrix.

Result Analysis

Transformed feature vectors are split into 80-20 training-testing respectively. Ratings are categorised into positive and negative denoted by 1 and 0. Three linear classification algorithms have been put into practice on five small samples created using K-fold from the training set. The accuracy, accuracy, recall f1 score of each sample was then calculated.

Algorithms	Accuracy	Precision	Recall
SVM	0.91	0.91	0.91
MNB	0.90	0.90	0.90
RandomForest	0.90	0.90	0.90

Table 1. Result analysis.

Conclusion

- Large amounts of feedback may be handled using this model, which also offers improved legitimacy.
- An aspect-based opinion miner in consumer domains that automatically identifies significant aspects and opinions of reviews, to create sentiment profiles.
- Our further study seeks to incorporate the aforementioned as well as the analysis of various sentence types, such as comparative and conditional sentences

References

[1] E. Cambria, B. Schuller, B. Liu, H. Wang, and C. Havasi, "Guest editorial special issue on concept-level opinion and sentiment analysis," IEEE Intelligent Systems Magazine, Special Issue on Concept-Level Opinion and Sentiment Analysis, vol. 28, pp. pp–15, 2012.

[2] A. Alghunaim, "A vector space approach for aspect-based sentiment analysis," Ph.D. dissertation, Massachusetts Institute of Technology, 2015.

[3] E. Cambria, B. Schuller, B. Liu, H. Wang, and C. Havasi, "Guest editorial special issue on concept-level opinion and sentiment analysis," IEEE Intelligent Systems Magazine, Special Issue on Concept-Level Opinion and Sentiment Analysis, vol. 28, pp. pp–15, 2012.

[4] B. Liu, "Sentiment analysis and opinion mining," Synthesis lectures on human language technologies, vol. 5, no. 1, pp. 1–167, 2012.