National College of Ireland

Classification of Customer Satisfaction for the Development of Hospitality Businesses

MSc Research Project
Data Analytics

Ruchira Talekar
Student ID: X18185703

School of Computing
National College of Ireland

Supervisor: Dr. Paul Stynes
Dr. Pramod Pathak

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Ruchira Talekar
Student ID:	X18185703
Programme:	MSc. Data Analytics **Year:** 2020
Module:	Research Project
Supervisor:	Dr. Paul Stynes, Dr. Pramod Pathak
Submission Due Date:	28th September 2020
Project Title:	Classification of Customer Satisfaction for the Development of Hospitality Businesses
Word Count:	7977 **Page Count:** 18

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Ruchira Talekar

Date: 28th September 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Classification of Customer Satisfaction for the Development of Hospitality Businesses

Ruchira Ramesh Talekar
x18185703

Abstract

Customer perception is helpful in the development of hospitality businesses and customer reviews are the source to know the satisfaction of the customer. Classification of customer satisfaction can help hospitality businesses like a restaurant for their improvement as per the customer's view and can improve negative services according to it. This research is based on big data handling and natural language processing. Research has analyzed the reviews of customers given for multiple hospitality businesses and classifying their sentiments into three classes, positive, negative, and neutral to predict customer satisfaction. The datasets used for the research are customer tweets using twitter streaming API and Yelp organization data, which contains various factors like, customer reviews, business name and id, states, ratings, etc. For the classification of sentiments, random forest, support vector machine and multinomial naïve Bayes model have been used. While k-means algorithms have been implemented to group similar customers together. As compared to naïve Bayes, the SVM algorithm has shown more accuracy. But the random forest is giving the best accuracy i.e. 90% by performing the job of feature selection. This classification can be helpful for business development and for making managerial decisions for customer satisfaction.

1 Introduction

Sentiment analysis of business reviews is facing various challenges in today's world. Sometimes, labeled data is insufficient while sometimes there is no capacity for good performance in various models. A complex sentence is one more challenge, as dealing with it is not the same as simple analyzing, because it needs more sentiments words. Singh et al. (2018) have used mine of data i.e. twitter for the analysis of consumer's tweets and predictions of negative reviews in supply chain management. Customers are preferring social media to show their feedback to a specific business. He et al. (2013) have designed a model for real-time data of the pizza industry using social media platforms like Twitter and Facebook.

This work is the following research by Singh et al. (2018) and Sánchez et al. (2019). This research is finding, up to what extent sentiment analysis of customer reviews given for various businesses can be used for the development of small businesses like restaurants, retail stores, parlors, salons, electronics, etc. This proposed research is based on classification customer reviews related to various businesses using an effective machine learning model. The data sets are used for this research are in JSON format, first one is customer tweets given on twitter (streaming API) while another is yelp organization data set. Independent columns are customer reviews in text format, city, states, user id, business id, rating given for the businesses while the dependent column is sentiments of the customers that have been created

by analyzing the data. SVM, Multinomial naïve Bayes and the random forest has been implemented to classify the customers' reviews into three categories, positive, negative, and neutral. K-means with TF-IDF clustering model has been designed to create the groups of similar reviews. Limitations of this research are twitter data that is extracted from twitter API is not that much suitable for this research as expected tweets i.e. customer reviews are unavailable. Another is, Yelp organization data is available is a large amount and it is the toughest task to split this data and load on the system. Preprocessing of text data is one of the major time-consuming tasks, that results in the accuracy of the models.

This research can help to improve the business as per the customer's needs. The clustering algorithm can be helpful to find the root cause of similar issues, while classification algorithms are helpful to analyze the sentiments of the customers. The relation between the negative reviews and drawbacks of the businesses is analyzed in this research and they are providing valuable insights for the growth of the businesses. Random forest algorithm has been designed to improve the accuracy of previous models and it is evaluating features selection from large data sets by showing the best performance. Businesses can improve their service by resolving the issues stated in negative reviews while analyzing positive reviews they can provide similar services to other customers for their satisfaction.

The research paper is structured in various sections given below: Section 2 is illustrating Related Work and Section 3 is describing the Research methodology. While Section 4 is showing the Design Specification of the research and Section 5 is explaining the Implementation flow of the research. Section 6 is an Evaluation, that contains various experiments and their performance measure as well as discussion of all the experiments. Last is Section 7 which is covering Conclusion and Future Work of the research.

2 Related Work

Various researchers Singh et al. (2018), Sánchez et al. (2019) have worked on business management, text data, and data analysis. For proposed research, different types of journals from the year 2011 to the year 2020 have studied. It includes 2 subsections, subsection 2.1 for business management and text analytics while another is subsection 2.2 for a summary of related work.

2.1 Study on Business Management and Text Analytics

Customer reviews are becoming a major factor in the development of various businesses. The source of these reviews can be found on social media, the reviews given on the websites of businesses, or can be given directly to the organization. This section is including the literature based on business management with the help of text analytics. Twitter dataset, Facebook dataset, Yelp datasets are included in it.

Singh et al. (2018) have researched the Beef Twitter dataset, which has been extracted using the twitter streaming API. They have implemented term frequency with support vector machine and naïve Bayes algorithms to predict issues of supply chain management using big data analytics. The support vector machine giving more better results as compared to the naïve Bayes model by showing 62.25% accuracy while naïve Bayes has been shown 39.71% i.e. very less accuracy. This research is based on Singh et al. (2018) as they have proved that twitter is a huge source of the dataset that shows the customer's opinion easily. Additionally,

they have used the n-gram features to improve the results of designed models. While Sánchez et al. (2019) has been designed model for the yelp dataset using a naïve Bayes algorithm. They have trained this model to classify the customer satisfaction level into positive and negative sentiments. They have worked on customer reviews based on Las Vegas hotels. The naïve Bayes has been shown 85% accuracy with a precision score of 0.89 for 47172 records. Sánchez et al. (2019) stated that 53% customers have avoided the services of the business that does not have reviews given by another customer.

He. et al. (2013) have been performed competitive analysis using various sources of datasets like Facebook and Twitter. Their research has been based on business management of the largest pizza chains by comparing the reviews given by the customers on Facebook and Twitter. They have implemented text mining and examined that the social media data is helping for the managerial decisions for the growth of the business by knowing customer's perceptions. It also gives an understanding of the competitors as well as the competition around the world. Ren et al. (2016) has used the traditional techniques of text mining with a recursive encoder to classify customer tweets. They have implemented Latent Dirichlet Allocation, logistic regression, and SVM for twitter user's sentiments classification. The logistic regression is showing an accuracy 78.57% and SVM is showing 77.44%. With the help of word embedding and by combining both the algorithms they have improved the accuracy to 81.02%. Xiong et al. (2018) have designed a model using a combination of lexicon resource and distant supervised information to classify twitter and word sentiments. It contains parallel symmetric ANN where MSWE is giving the highest accuracy amongst random, Word2Vec, C&W, and SSWE by giving 85.75% accuracy. To improve the performance of the model they have used the unigram, bigram, and trigram features with MSWE. Zhang et al. (2019) has proposed the 3W-CNN model using traditional text mining for sentence-level sentiment classification. They have compared the combination of naïve Bayes and SVM with 3W-CNN while deep learning has been shown the best performance by showing 90.3% accuracy. Alharbi et al. (2019) has implemented deep neural network CNN to classify the twitter sentiments into positive, negative, and neutral polarity using customer's behavioral information. They have compared this model with SVM, Naïve Bayes, KNN, and LSTM, while CNN has been given 0.88 precision and 0.89 recall. Spelling errors, tweets length, abbreviations, special character are some cons of twitter data and hence it needs a non-traditional approach like CNN. They have analyzed customer's attitude and social activities to improve the business issues. Saif et al. (2012) have proposed the model for the prediction of people sentiment using three different datasets. They have segregated the correlation between sentiments and products purchased by the customer by adding semantics. It is giving 83% accuracy while unigram has been shown less accuracy and semantics has been improved the accuracy. The author has conveyed that as per each dataset accuracy of the models is changing. The positive classification has been shown more precision while negative classification has been represented greater F-score and recall. Ghiassi et al. (2013) have designed dynamic ANN with n-gram features for the sentiment analysis of big data based on social media. They have used twitter lexicons to improve corpus coverage. It is giving 96% accuracy for positive class while 89% accuracy for strongly negative class and 50% for mildly negative class with respect to 10 million records of the dataset.

F.F. et al. (2014) have developed the model using lexicons and machine learning classifiers for the classification of customer reviews given on twitter automatically. Bag of words has been used with four different models NB, Logistic Regression, Random forest to compare the results. The ensemble model has been given 70.80% to 87.20% accuracy for various five datasets. Ankit et al. (2018) have designed the ensemble classifier that combing the performance of the multiple base classifiers into a single unit to classify the consumer tweets and to predict their satisfaction level. NB, Logistic Regression, Random forest, SVM

has been given low accuracy when they were standalone while the ensemble model has been shown greater i.e. 73% to 85% accuracy for different four datasets. Ruz et al. (2020) have been implemented the Bayesian network for the classification of natural disasters. They have used Spanish datasets of the Chilean earthquake and to improve the accuracy Bayes factor has been applied with the Bayesian network. It has been compared with random forest and SVM. SVM has given accuracy by showing 81% and the random forest has given 85% for another dataset, while it has been analyzed the historical and social information of the events.

Rafeeque et al. (2015) have been designed a model using the unsupervised approach and lexicons to analyze the customer's sentiments. Sentiments lexicons SentislangNet, SenticNet, and SentiWordNet have been used for the evaluation of the tweets posted by customers. SentiWordNet and SenticNet have been combined to build the SentislangNet. These lexicons have been helped to increase the accuracy of the models for the predictions of the sentiments. Parallel python has been used to handle large scale data and it has been shown accuracy 63%. Hartmann et al. (2019) have researched automated five lexicon and five classifiers of machine learning using four social media text datasets of various languages. LIWC (linguistic enquire), as well as word count, has performed poorly than machine learning classifiers (NB, Random Forest, KNN, ANN, SVM). The Naïve Bayes has been shown the best accuracy for small size datasets (75.4%) while random forest has been shown best accuracy for large size datasets (70%). Sailunaz et al. (2019) have implemented a model using sentiment scores to analyze the emotions of the customers using social media data set. Naïve Bayes has given 66% accuracy amongst all the models like Random Forest, SVM that have been executed for emotions and sentiments. Bai, X., (2011) has been implemented a prediction system using a heuristic search with a Markov blanket model for sentiment analysis of customer. Movie reviews and online news datasets have been used while the model has been shown 95% accuracy. Chang et al. (2019) have been used visual analytics to classify the reviews of Tripadvisor dataset. They have observed that business travelers have been shown a low rate as compared to couples travelers using a pie chart, geographical analysis, word cloud, timeline analysis, etc.

Herbst et al. (2020) have been researched on childcare businesses and evaluation of satisfaction of parents using the Yelp consumer reviews dataset. They have implemented regression techniques that showed 80% accuracy. While Nakayama et al. (2018) has been developed a model to classify Yelp reviews based on Japanese restaurants for sentiment analysis. They have observed the difference in the culture of Japanese and Western consumers by 72% for the quality of food and more than 100% for service quality. Li Q. et al. (2020) have been proposed a text classification model using improved CNN with the help of recursive data trimming for addressing problems that are not fitted like irrelevant words. The model has been developed for Yelp reviews dataset and it has been given improved 85% accuracy. Liu et al. (2019) have been explored the language style to know the quality of reviews given by the customers. The yelp customer review has been used for language style matching and electronic word of mouth to improve the results. Banerjee et al. (2017) has developed the Logistic Regression model for identification of customer's trustworthiness by analyzing the Yelp review dataset. The author stated that the trust of a customer is highly impacting product sales and the logistic model has been shown 83% accuracy for it. Jimenez et al. (2019) have been developed two stages framework for text analytics of social media content. They have analyzed small and big Yelp reviews datasets, while Multi-Layer Perceptron with TFIDF vectorizer has been shown best accuracy i.e. 88% for binary classification.

Nguyen et al. (2019) have proposed the ensemble method with k-means and Birch clustering for the sentiment analysis of the twitter dataset. The CNN and LSTM models have been developed for the NLP on tweets using five datasets of movie reviews that have been

shown 88% accuracy together. For improving the weakness of the ensemble model, they have added a clustering strategy that has been used for the grouping of semantic similarities. Singh et al. (2018) have been developed hierarchical clustering with classification models to find the similarities between the customers and the root cause of the dissatisfied customers of supply chain management. This hierarchical clustering has been improved the accuracy using multiscale bootstrap resampling that creates clusters for informing about drawbacks mention by the customers.

2.2 Summary

This research is based on the research work of Singh et al. (2018) and Sánchez et al. (2019). Table 1 is illustrating the summary of both the research work. The limitation of Singh et al. (2018) is research has been limited to the beef twitter dataset. They have implemented SVM and Naïve Bayes models as well as hierarchical clustering, but it has shown less accuracy for word features due to inappropriate data cleaning and pre-processing. This research has been worked on lamb and pork dataset and overcomes the limitations of the above researchers by cleaning and pre-processing the data properly. While Sánchez et al. (2019) has been developed the Naïve Bayes model for Yelp customer reviews dataset and their limitation is, they have worked on only Urban hotel reviews of Las Vegas city. This research has included various types of business reviews of the Yelp dataset so that research is not bias and hence it is improving the previous work. Current research has been implemented a random forest algorithm as well as previous work has been improved by using n-gram and TF-IDF features. Additionally, the k-means clustering algorithm has been implemented for the grouping of a similar customer. Sentiment analysis and clustering can be helpful for the managerial decisions of the business team.

Table 1: Related Work Summary

Author	**Method used**	**Dataset**
Singh et al. (2018)	SVM, Naive Bayes	Beef Twitter Dataset
Sánchez et al. (2019)	Naïve Bayes	Yelp Customer Reviews Dataset

3 Research Methodology

This section of the research project is explaining multiple steps that have been carried out during the research. It contains data gathering of Twitter and Yelp data set, data processing of tweets and customer reviews, classification of reviews using various machine learning models, clustering to create a grouping of similar reviews and evaluation process to know the performance of the models. Figure 1 is representing the flow of research methodology.

3.1 Data Gathering using Twitter API and Yelp Organization Dataset.

In this research, data has been collected from two ethical and open source websites. Twitter data set has been retrieved from twitter streaming API[1] by creating a new account and using consumer and API tokens provided by twitter. Tweepy is a python library that has been used to retrieve the data and perform various tasks. This data is in JSON format. 90000 lamb and pork business reviews in the form of tweets have been extracted using the feature keywords. Singh et al. (2018) have used beef reviews in the form of tweets and it has been extracted from twitter API. The Independent variable of this data set is twitter text, location, created date, keyword, hashtag while the dependent variable is created further as the sentiment of the tweets. While another data set that has been used for this research is the Yelp organization dataset[2] based on customer reviews given for various businesses. This data set is available on the Kaggle website and it is open source.

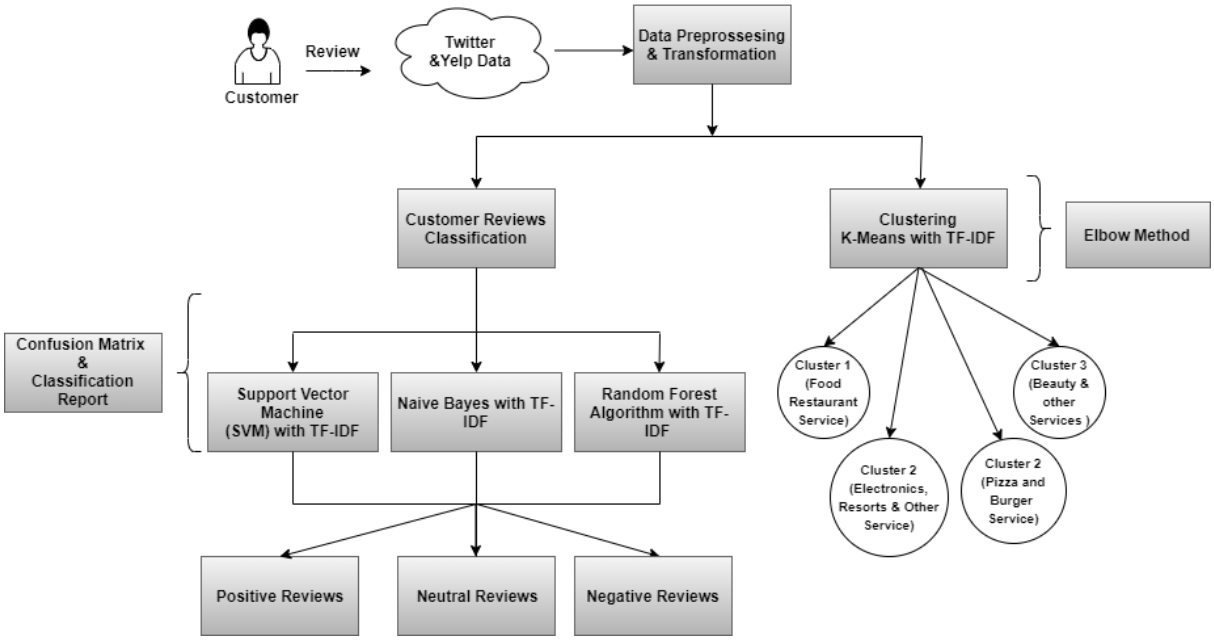

Figure 1: The Flow of Research Methodology

The independent variables of the data set are, review text, city, user id, business id, star ratings, created date, and review id. The dependent variable has been created further as a sentiment of review, which is explained further.

3.2 Data Selection, Pre-processing, and Cleaning.

From twitter API, 90000 records have been extracted ethically, but 80000 tweets have been selected by removing unrelated tweets. While for the Yelp data set, 130000 rows of data have been selected from the huge amount of data. For data pre-processing, JSON data has been converted into a structured tabular format for both the data sets. Then removed the missing data. For twitter data set, Tweets Sentiments as a target variable has been created by using

[1] https://developer.twitter.com/en
[2] https://www.kaggle.com/yelp-dataset/yelp-dataset

textblob library and it has factors positive sentiment and negative sentiment. While for the yelp data set, the star rating column has been used to create target column review sentiments. This target variable has been created by, 1 and 2 ratings are observed as a negative review, 4 and 5 ratings are observed as a positive review and 3-star rating has mixed reviews, hence it is given as a neutral review. Furthermore, both data sets have been cleaned using various packages. For review text, it has been removed to stop words, punctuation, HTML tags, numbers, white space as well as converted letters to lower case and applied to stem. Then words and sentences have been transformed with the help of lemmatization and tokenization using appropriate packages.

3.3 Data Transformation using Vectorization and Document Term Matrix.

Data transformation has followed various types of processing on text data. The term frequency-inverse document frequency feature has been used to check the importance of each word in a corpus. It is helpful to determine the frequency of the word that appeared in a customer review. Then, a bag of words has been created and it is also called a document term matrix. This document term matrix has been performed as an important task by representing the count of each word into each document. Each word is termed as a feature. In this research, the training vector has (64000, 20144) features.

3.4 Classification of the customer reviews using Support Vector Machine and Naïve Bayes.

Classification of customer reviews for both the data sets has been implemented using the support vector machine and to compare the results it has been also classified using the naïve Bayes model. For both the models, training data is 80% while testing data has been 20%.
For the Twitter data set, the target variable has two factors as positive and negative tweets, so that model has been trained and data is classified into positive and negative sentiments of the customers. For Yelp organization data, the target variable has three attributes, positive review, negative review, and neutral review. Both models have been implemented to classify customer reviews for various businesses. To improve the accuracy of these models TF-IDF has been implemented. This experiment has been performed to replicate previous experiments by Singh et al. (2018) and Sánchez et al. (2019).

3.5 Classification of customer reviews using the Random Forest Algorithm.

Random forest algorithm is useful for the classification of data as it gives the results by averaging the results of multiple decision trees. Customer reviews of Yelp data set have been classified using by implementing random forest algorithm with TF-IDF and classified data into the positive review and negative review as well as neutral review. Mostly this experiment has been focused on positive and negative sentiments to improve the accuracy of the model. Furthermore, unigram and bigram features have been combined with the model to increase the performance of the model. Also, it is finished executing within minimum time than support vector machine and naïve Bayes.

3.6 Evaluated Classification Algorithms using Accuracy, Precision and Recall

To check the performance of the SVM, naïve Bayes, and random forest algorithm, models have been tested using 20% of the data set, and the accuracy, precision, and recall have been measured. The confusion matrix has been showing the results for it. SVM is giving accuracy 94% while multinomial naïve Bayes is giving 84% accuracy for twitter data classification. Next, these models are giving 85% accuracy for and 74% for yelp organization reviews classification using SVM and naïve Bayes respectively. While, for the random forest algorithm it is giving accuracy 78% and 90% for the model without unigram, bigram, and model with unigram, bigram, respectively.

3.7 Finding of number of clusters using elbow methods and grouping of similar customers feedback based on various cities using k-means clustering with TF-IDF

Furthermore, the elbow method has been implemented to find out the appropriate number of clusters for the k-means clustering algorithm with TF-IDF vectorization. The elbow point is at 4 and hence, k-means have been designed for four number of clusters. It has been shown the specific results by grouping similar customer reviews into the same cluster by their reviews and various cities.

4 Design Specification

The framework contains the overall flow of this research for the prediction of customer satisfaction with the help of clustering and classification. Figure 2 is representing the pictorial view of the framework.

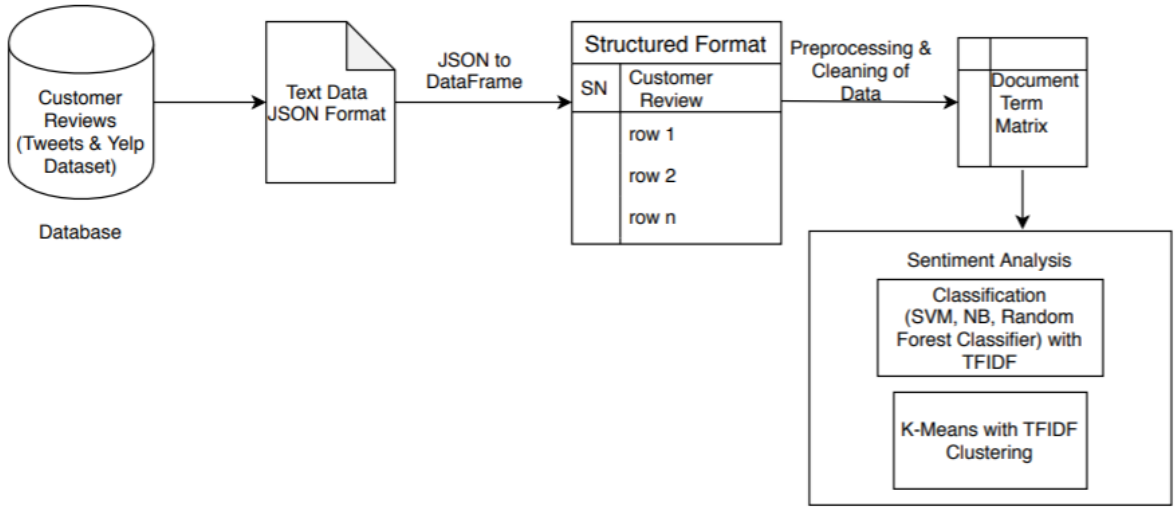

Figure 2: Framework for Classification of Customer Satisfaction

The first phase is a database, that contains multiple datasets which also includes the twitter dataset and the Yelp dataset. However, in the next phase, the twitter dataset and Yelp dataset has been extracted from the database in JSON unstructured format. Furthermore, unstructured

text data has been converted to structured DataFrame. The dataset in tabular format has been shown columns like customer reviews, rating, city, sentiments, etc. In the next phase, pre-processing and cleaning of the datasets have been performed and data has been transformed into Document Term Matrix (DTM). In the last phase, sentiment analysis with the help of classification and clustering algorithms has been performed.

5 Implementation

The implementation of this research has been performed using Python programming language and Jupyter Notebook as a programming tool. Twitter dataset and Yelp dataset has been pre-processed and transformed using various python libraries. Twitter dataset and yelp dataset has been in unstructured JSON format and it has been converted into a structured format. Furthermore, implementation is performed for data transformation, document term matrix and vectorization, to develop the models like SVM, Naïve Bayes, random forest and k-means clustering.

5.1 Data Transformation and Tokenization

The major task was cleaning of text data by removing unwanted things. NLTK library has been used to perform various processing on customer reviews. Re library has been used to detect the various punctuation, HTML tags, number, usernames and removed it to clean the data. These reviews have been converted into the lower case by using lower command and removed white space by using whitespace command. Then tokenization has been performed to transform the data using the NLTK library. Lemmatization and stemming of data have been implemented using NLTK and Snowball Stemmer library. It has transformed the sentences and words into small tokens that have suitable for further processes like corpus and DTM. NLTK library feature stopwords have been used to detect it, remove it using remove () command, and store the new data in a corpus. It is useful to remove the words like is, are, to, no, I, we, me, you, etc.

5.2 Document Term Matrix and TF-IDF

The term frequency-inverse document frequency feature has been used to check the importance of each word in a corpus. It has been used to determine the frequency of the word that appeared in a customer review. Then, the document term matrix has been created to represent the count of each word in each document. To perform TF-IDF and vectorization, sklearn library has been used and it contains feature TfidfVectorizer. For this research, the size of the created vector for the twitter data set is (37000, 20000), while the size of the created vector is (64000, 20144) for the yelp dataset.

5.3 Naïve Bayes and Support Vector Machine Classifier

The next phase of implementation is to design and implement the various models that are applicable to get solution for the research questions. For both the models, training data is 80% while testing data has been 20%. In this research, for comparison of performance three classifiers that have been applied to classify the customer reviews and to determine the sentiments of the customers as well as their satisfaction level. Multinomial Naïve Bayes first

model that has been designed to classify the reviews and tweets into various levels of sentiments. It is a supervised machine learning model. The independent variable of the twitter dataset is tweets posted by the customer for beef and pork business while the dependent variable is sentiments of the tweets that have been created using textblob library. For yelp organization data set independent variable is customer reviews for various businesses and the dependent variable is a sentiment of the customer that has been created based on ratings given by the customer. The second model that has been implemented is the support vector machine and it is an unsupervised machine learning model. The kernel used for SVM is a linear kernel, to classify the reviews and tweets into positive, negative, and neutral sentiments. The third model that has been designed is random forest algorithms. Both models have been performed using sklearn library.

5.4 Random Forest Classifier

Naïve Bayes and SVM models have been implemented for the twitter dataset and Yelp dataset, while the random forest algorithm has been implemented only for the yelp data set. Training data has been 80% while testing data has been 20% for the model. Sklearn and pickle library has been used for the implementation of the random forest classifier. A document term matrix has been created for this model. Amongst all these models, the random forest algorithm is showing the best performance to classify the customer reviews into positive and negative sentiments. To improve the accuracy of the random forest model n-gram with TF-IDF features have been applied with the model. Unigram and bigram features have been applied with the random forest algorithm that has been improved the accuracy of the model. The random forest classifier has been completed it's training and testing in minimum time as compared to other experiments.

5.5 K-means Clustering with TF-IDF

The last phase of implementation has been covered the grouping of similar customers to find out the solution of various problems. Transformed data with TF-IDF vectorization has been used as input data for the clustering. The elbow method has been implemented to know the perfect number of clusters for the Yelp dataset. Matplotlib and sklearn.cluster library has been used for it. Elbow method has shown the elbow point at 4. K-means clustering has been implemented using the number of clusters is equal to four. It has given the four clusters segregated with restaurants, pizza, and burgers shops, beauty parlor and other services, electronics services. It has taken maximum time for the evaluation as well as for implementation, as compared to other experiments.

6 Evaluation

After research implementation, it is necessary to know the performance of the models that have implemented. In this research, four experiments have performed. The first three experiments are implemented, to classify the customer reviews for prediction of customer satisfaction using Support Vector Machine, Naïve Bayes, and Random Forest classifier. They are evaluated using confusion matrix, accuracy, precision, and recall. The fourth experiment

is based on k-means clustering with TF-IDF to find the similarities of customers. It is evaluated using the elbow method.

6.1 Experiment 1: Sentiment Analysis using SVM and Naïve Bayes for Twitter Dataset

Experiment 1 is the classification of customer tweets of lamb and pork using support vector machine and naïve Bayes algorithms. These models have been classified as the text into positive and negative classes. Positive reviews are stating satisfied customer, while negative reviews are stating dissatisfied customer.

This experiment has been implemented to replicate the research of Singh et al (2018). The support vector machine has been designed with the help of the TF-IDF feature. Customer reviews in terms of tweets have been classified and shown the accuracy of the model 94%. This model has improved the problem using the TF-IDF feature and Document Term Matrix as compared to the SVM model of Singh et al (2018).

Another model of experiment 1 is Naïve Bayes. It has classified lamb and pork tweets into two classes positive and negative using feature TF-IDF vectorization and bag of words. It has shown the precision and accuracy of 84%. It has shown less performance as compared to SVM, but better than Singh et al (2018). Pre-processing and cleaning of text data are always helping to increase the performance of the model which is lacking in the experiments of Singh et al (2018). A comparison of the performance of SVM and Naïve Bayes algorithms has shown in Figure 3.

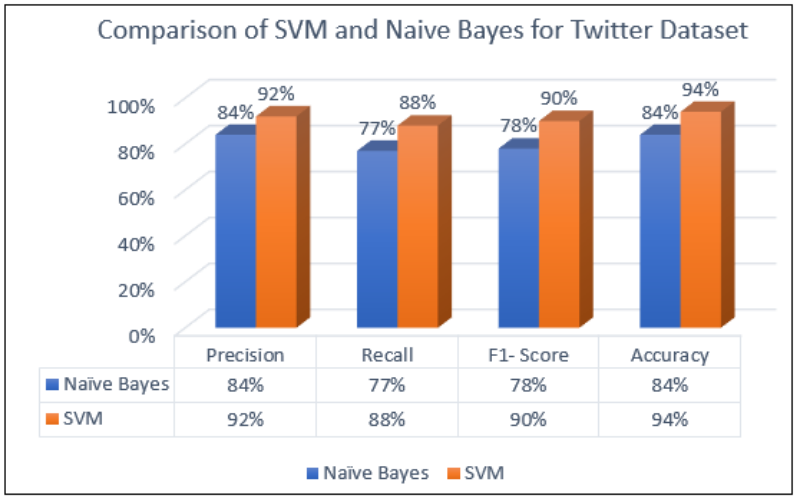

Figure 3: Classification Report for SVM and Naïve Bayes for Twitter Dataset

Analysis of twitter dataset and observations of results has concluded that twitter data is not providing valuable insights as customer tweets are mostly not related to the supply chain business of lamb and pork. Hence, the next experiment has been performed with another dataset.

6.2 Experiment 2: Sentiment Analysis using SVM and Naïve Bayes for the Yelp Dataset

In experiment 2, the data set has changed as the twitter data set is not containing good insights related to customer reviews. Hence, for the second experiment Yelp organization data set has been selected. For comparison with Singh et al (2018) and Sánchez et al. (2019), support vector machine, and Naïve Bayes models have been designed for classification of customer reviews.

As compared to previous results SVM of Singh et al (2018), accuracy has been increased by 23%. Experiment 2 is showing 85% accuracy for support vector machine with TF-IDF word features TF-IDF. Multinomial naïve Bayes is showing 74% for the Yelp dataset, but it is less than Sánchez et al. (2019). Henceforth, it is concluded that the support vector machine is showing the best performance as compared to multinomial naïve Bayes. For better understanding, Table 2 is representing the classification report for SVM and naïve Bayes for the Yelp dataset.

Both algorithms have not performed the task of feature selection. Hence, to improve the accuracy and to select the more effective classification model, the next experiment has been performed.

Table 2. Classification Report of Naïve Bayes & SVM for Yelp Dataset

Model Names ->	Naïve Bayes		SVM	
Sentiments ->	Positive	Negative	Positive	Negative
Precision	0.78	0.70	0.89	0.81
Recall	0.85	0.73	0.95	0.84
F1- Score	0.81	0.77	0.92	0.82
Accuracy	74%		85%	

6.3 Experiment 3: Classification of Customer Reviews using Random Forest for the Yelp Dataset

In experiment 3, the Yelp organization data set has opted for classification of customer reviews into three classes positive reviews, negative reviews, neutral reviews using the document term matrix, and random forest algorithm. As compared to previous results of Singh et al (2018) and Sánchez et al. (2019), accuracy has been increased and it is giving the best performance as compared to experiment 2. Table 3 is representing a classification report for various experiments of the random forest model.

Table 3. Classification Report of Random Forest Model

Model Names ->	Random Forest with Neutral			Random Forest without Neutral		Random Forest (n-gram)	
Sentiments ->	Positive	Negative	Neutral	Positive	Negative	Positive	Negative
Precision	0.82	0.72	0.35	0.91	0.81	0.92	0.83
Recall	0.94	0.68	0.09	0.94	0.74	0.95	0.77
F1- Score	0.87	0.7	0.14	0.93	0.78	0.93	0.8
Accuracy	78%			89%		90%	

For experiment 3, various models have been designed using a random forest algorithm. One of the models is customer reviews have been classified using the above three classes and it has shown accuracy 78%. But, if only positive and negative classes are selected for classification, then it has shown an accuracy of 89%. By implementing TF-IDF vectorization and unigram as well as bigram, the random forest model has given the highest performance i.e. 90%. This experiment is helpful to show results using various features of text analytics. The confusion matrix is, the true positive count is 3569, the false-positive count is 1064, the false-negative count is 712, the true negative count is 12425. Figure 4 and Figure 5 are illustrating the word cloud of the classified positive and negative reviews.

Classifiers have classified the reviews based on sentiments of the customer. To overcome this limitation, the next experiment has been performed where customers are grouped as per the similarities of reviews.

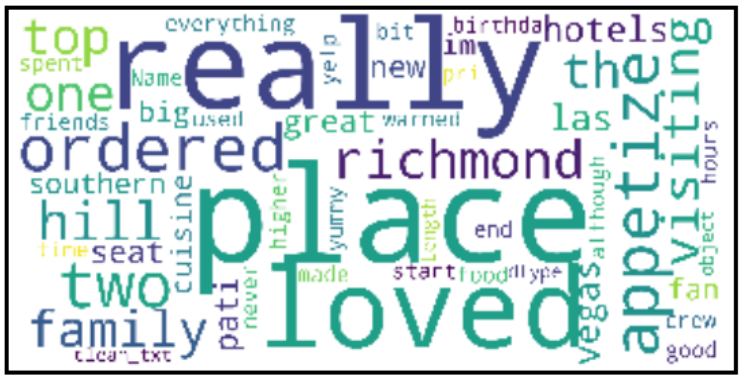

Figure 4: Word Cloud of Positive Customer Reviews

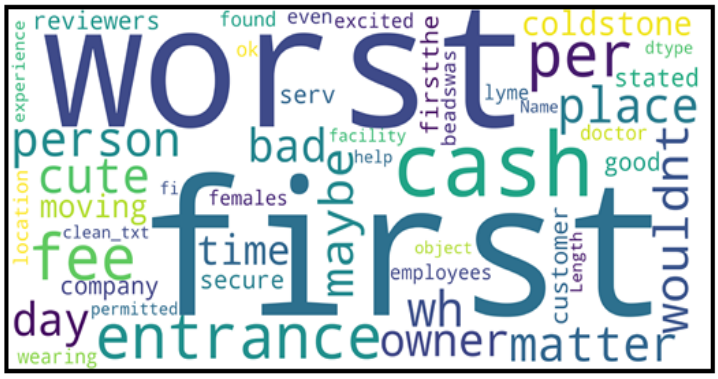

Figure 5: Word Cloud of Negative Customer Reviews

6.4 Experiment 4: K-means Clustering with TF-IDF for Yelp Dataset

In experiment 4, as per different similarities, customer reviews are grouped according to various cities and implemented, k-means clustering algorithm, with TF-IDF vectorization. Singh et al (2018) have performed clustering and this experiment has improved their experiment. Elbow method has been applied to find the perfect number of clusters for overall data. As per the elbow method, Figure 6 is illustrating the elbow point 4 is suitable for the

number of clusters. Furthermore, by taking the four clusters, the clustering algorithm has been implemented. These clusters are helpful to know the similarity between the customer's opinion as per different cities and various hospitality services. The 1st cluster is containing reviews related to food restaurant and it contains 19577 reviews. The second cluster is containing reviews related to resorts and electronics services business while its count is 15837. The third cluster is containing reviews related to beauty and other services reviews and the total review count is 11932. The fourth cluster has 16654 reviews about pizza and burger shops. K-means Clustering with TF-IDF vectorization has been more suitable for big data as compared to hierarchical clustering.

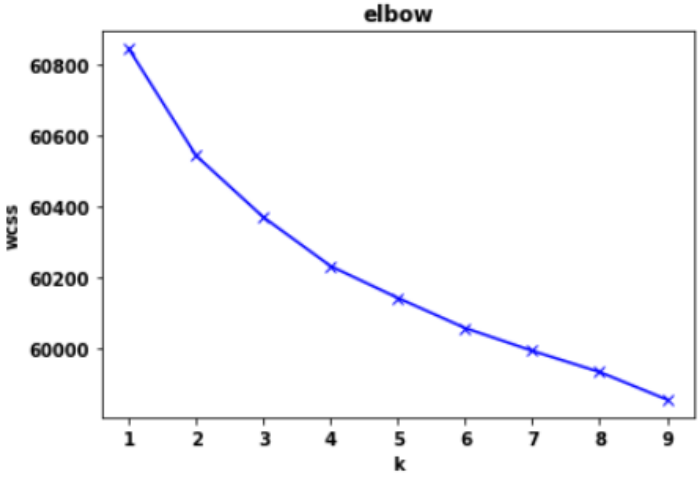

Figure 6: Elbow Method

6.5 Discussion

The goal of this research was to build a model that can successfully classify customer reviews to predict their satisfaction level and for the development of hospitality businesses. Experiment 1 was a classification of lamb and pork tweets using term frequency with support vector machine and naïve Bayes classifiers. This experiment has been performed to compare the results with Singh et al. (2018). Results are aggregable with Singh et al. (2018) as the SVM model has shown the best performance as compared to naïve Bayes. The accuracy of the SVM is 94% and naïve Bayes is 84%. But, for the lamb and pork dataset, they have shown 13% and 23% better performance for SVM and Naïve Bayes respectively, as compared to Singh et al. (2018). To improve the limitations of Singh et al. (2018), the dataset has been pre-processed and cleaned using more features that were helpful for the better performance of experiment 1. But, after analyzing the results of Experiment 1, it has been observed that twitter data does not have meaningful data for business analysis. Tweets in terms of customer reviews are unavailable in this data and instead, it contains other types of reviews and comments. To overcome this observation, in a second experiment new dataset has been selected.

In experiment 2, support vector machine and multinomial Naïve Bayes with term frequency have been implemented for the classification of customer reviews using the Yelp organization dataset. The SVM has been shown 85% accuracy while multinomial naïve Bayes has been shown 79% accuracy. It is better than previous experiments Singh et al. (2018), but naïve Bayes model results are 6% less than Sánchez et al. (2019). Additionally, it is observed that, for these models, instead of performing with multiple values, the dependent variable with binary class has shown much better results. As per the analysis, Yelp datasets

reviews (Experiment 2) has given more valuable insights as compared to twitter dataset (Experiment 1). To increase the accuracy of Experiment 2 using feature selection, the next experiment has been implemented.

The third experiment was performed for the classification of customer reviews using a random forest model with the various feature set. The random forest model has been designed only for the Yelp dataset. Firstly implemented, random forest algorithm using target column values positive, negative, and neutral, while it has shown accuracy 78%. Secondly, the random forest model built with binary target column values positive, negative that has shown accuracy 89%. To improve this result, the random forest algorithm is trained with a combination of term frequency-inverse document frequency and unigram, bigram. This experiment has given the highest results by showing a 90% accuracy. This experiment has shown 5% better accuracy than the naïve Bayes model of Sánchez et al. (2019). The comparison between machine learning classifiers that have implemented in this research shown in Figure 7.

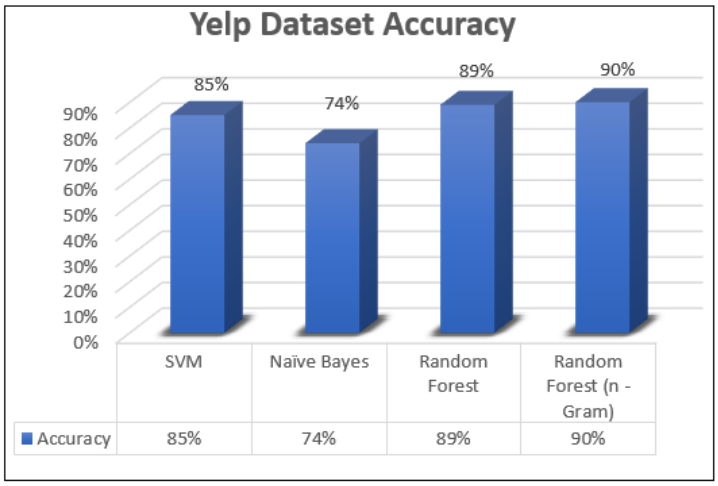

Figure 7: Comparison of Accuracy of Classifiers

In experiment 4, to find out the similarities between the customers, clustering has been implemented. As per the elbow method, K-means clustering with the TF-IDF model has been designed with four clusters. The first cluster is showing similar customers of food restaurants by grouping their reviews. The second cluster is representing the resort and electronics services customers. The third cluster is containing customers of beauty and other services reviews. While fourth is containing reviews of pizza and burger shops. This research is helpful to know the key features that are noticed by the customers for the positive or negative evaluation of the services. Reasons for customer satisfaction have been shown by clustering and they are like food taste, quick service to operate the electronics things, pleasant experience at the resort due to various reasons, providing the best waiter and waitress service, family restaurants, etc.

The limitations of this research are poor insights from twitter dataset, pre-processing of a large amount of text data, Yelp organization data is difficult to manage due to the large quantity and the k-means model takes maximum time for the execution and evaluation using elbow method and vectorization.

The classification of customer reviews is predicting the sentiments of the customers for various businesses like lamb and pork supply chain, restaurant, resort, electronics services. This research is helpful for the managerial decisions of businesses by knowing the satisfaction level of customers. Classified positive and negative reviews are helpful to know the perceptions of the customers and it is helpful to improve the poor services of the

hospitality business with the help of negative reviews as per the customer's view. At the end of this discussion, the research work has contributed in the data analytics field by improving previous experiments, by implementing random forest algorithm with TF-IDF and n-gram features and by designing k-means clustering with TF-IDF vectorization.

7 Conclusion and Future Work

This research is to gain an understanding of how customer reviews can be used for the prediction of customer satisfaction and the development of hospitality businesses. Classification of the customer reviews has been implemented successfully using the twitter dataset and Yelp organization dataset. While twitter dataset is not giving the valuable insights that are useful as per the business perspective and hence, the Yelp dataset has been selected for better results. Due to the large quantity of text data, pre-processing and data cleaning has been time-consuming processes. Support vector machine, multinomial naïve Bayes, and random forest classifiers have been trained for the classification of the reviews. Additionally, n-gram features, unigram and bigram have been used to obtain more improved results. As compared to naïve Bayes, SVM is showing better results. Random forest algorithm with TF-IDF and n-gram features has shown the highest accuracy and reduced complexity. Hence it is selected as an effective customer reviews classification algorithm. However, K-means clustering with TF-IDF has been performed for the grouping of similar issues of customers. This research is helpful to know the key features that are noticed by the customers for the positive or negative evaluation of the hospitality services.

The results of this research are useful for future work. Deep learning method Convolutional Neural Network can be used for better feature selection to improve the results of the Yelp dataset. While for big data storage and handling of big data, cloud storage, and cloud services like AWS and Microsoft Azure can be used. For more understanding of similar customers, latent Dirichlet Allocation can be used.

Acknowledgment

I would like to thank my supervisors, Dr. Paul Stynes and Dr. Pramod Pathak for their support during this research study. I have achieved these results due to their continuous feedback and guidance in the right direction. I am also thankful to respective organizations for providing the open-source datasets and software for academic research work. Lastly, I would like to thank my parents and friends for their patience, support, and guidance throughout the research and course.

References

Alharbi, A. S. M. and Elise (2019). Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information, Cognitive Systems Research 54: 50 – 61.

Ankit and Saleena, N. (2018). An ensemble classification system for twitter sentiment analysis, Procedia Computer Science 132: 937 – 946. International Conference on Computational Intelligence and Data Science.

Bai, X. (2011). Predicting consumer sentiments from online text, Decision Support Systems 50(4): 732 – 742. Enterprise Risk and Security Management: Data, Text and Web Mining.

Banerjee, S., Bhattacharyya, S. and Bose, I., 2017. Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business. Decision Support Systems, 96, pp.17-26.

Chang, Y.C., Ku, C.H. and Chen, C.H., 2019. Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. International Journal of Information Management, 48, pp.263-279.

F.F., N., Hruschka, E. R. and Hruschka, E. R. (2014). Tweet sentiment analysis with classifier ensembles, Decision Support Systems 66: 170 – 179.

Ghiassi, M., Skinner, J. and Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network, Expert Systems with Applications 40(16): 6266 – 6282.

Hartmann, J., Huppertz, J., Schamp, C. and Heitmann, M. (2019). Comparing automated text classification methods, International Journal of Research in Marketing 36(1): 20 – 38.

He, W., Zha, S. and Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry, International Journal of Information Management 33(3): 464 – 472.

Herbst, C.M., Desouza, K.C., Al-Ashri, S., Kandala, S.S., Khullar, M. and Bajaj, V., 2020. What do parents value in a child care provider? Evidence from Yelp consumer reviews. Early Childhood Research Quarterly, 51, pp.288-306.

Jimenez-Marquez, J.L., Gonzalez-Carrasco, I., Lopez-Cuadrado, J.L. and Ruiz-Mezcua, B., 2019. Towards a big data framework for analyzing social media content. International Journal of Information Management, 44, pp.1-12.

Liu, A.X., Xie, Y. and Zhang, J., 2019. It's Not Just What You Say, But How You Say It: The Effect of Language Style Matching on Perceived Quality of Consumer Reviews. Journal of Interactive Marketing, 46, pp.70-86.

Li, Q., Li, P., Mao, K. and Yat-Man, E.L., 2020. Improving Convolutional Neural Network for Text Classification by Recursive Data Pruning. Neurocomputing.

Nakayama, M. and Wan, Y., 2018. Is culture of origin associated with more expressions? An analysis of Yelp reviews on Japanese restaurants. Tourism Management, 66, pp.329-338.

Nguyen, H. T. and Nguyen, M. L. (2019). An ensemble method with sentiment features and clustering support, Neurocomputing 370: 155 – 165.

Rafeeque, P., Selvaraju, S. and Mahalakshmi, G. S. (2015). Twitter sentiment analysis for large-scale data: An unsupervised approach, Cognitive Computation 7: 254–262.

Ren, Y., Wang, R. and Ji, D. (2016). A topic-enhanced word embedding for twitter sentiment classification, Information Sciences 369: 188 – 198.

Ruz, G. A., Henr´ıquez, P. A. and Mascaren˜o, A. (2020). Sentiment analysis of twitter data during critical events through bayesian networks classifiers, Future Generation Computer Systems 106: 92 – 104.

Saif, H., He, Y. and Alani, H. (2012). Semantic sentiment analysis of twitter, Lecture Notes in Computer Science 7649: 508–524.

Sailunaz, K. and Alhajj, R. (2019). Emotion and sentiment analysis from twitter text, Journal of Computational Science 36: 101003.

Sánchez-Franco, M.J., Navarro-García, A. and Rondán-Cataluña, F.J., 2019. A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services. Journal of Business Research, 101, pp.499-506.

Singh, A., Shukla, N. and Mishra, N. (2018). Social media data analytics to improve supply chain management in food industries, Transportation Research Part E: Logistics and Transportation Review 114: 398 – 415.

Xiong, S., Lv, H., Zhao, W. and Ji, D. (2018). Towards twitter sentiment classification by multi-level sentiment-enriched word embeddings, Neurocomputing 275: 2459 – 2466.

Zhang, Y., Zhang, Z., Miao, D. and Wang, J. (2019). Three-way enhanced convolutional neural networks for sentence-level sentiment classification, Information Sciences 477: 55 – 64.