# A Comparative Analysis of Text Mining Methodologies for Online Consumer Reviews

**Sonya Zhang**
**California State Polytechnic University, Pomona**

**Daniela Lopez**
**California State Polytechnic University, Pomona**

**Lyndon Lam**
**California State Polytechnic University, Pomona**

**Abhishek Yenumula**
**California State Polytechnic University, Pomona**

**Tony Vu**
**California State Polytechnic University, Pomona**

*Extracting meaningful insights from the sheer volume of Online Consumer Reviews (OCRs) has been challenging. We aim to explore the most effective methodologies for text mining of OCRs, covering topic extraction, topic classification, and sentiment analysis. Through a comprehensive review of recent research on text mining applied to OCRs, we found that LDA2Vec can enhance the effectiveness of conventional LDA for topic extraction. Additionally, the combination of Convolutional Neural Networks (CNN) and GloVe demonstrates the best performance for topic classification, while CNN and SVM outperform other algorithms for sentiment analysis. Furthermore, the spaCy Natural Language Processing (NLP) proves to be a more effective choice for text pre-processing compared to Natural Language Toolkit (NLTK). Subsequently, we applied these refined models to a Yelp reviews dataset, assessed their performance against conventional models, and provided a comprehensive discussion of the results and limitations. The insights gained from this study can be valuable for developing effective models in OCR analysis.*

*Keywords: online consumer reviews, Yelp, machine learning, natural language processing, topic modeling, LDA, sentiment analysis, text classification, CNN, Word2vec, GloVe*

## INTRODUCTION

Online customer review platforms like Yelp or TripAdvisor have grown significantly in popularity and have become an essential tool for both consumers and businesses across hospitality and restaurant industries and many others. These platforms provide an opportunity for consumers to share their experiences with others, and businesses can use the feedback to improve their products or services. Consumers often rely heavily on reviews to inform their purchasing decisions, leading to a significant impact on various aspects of a business, such as sales (Ye et al., 2009). At the same time, businesses can benefit from consumer feedback to identify areas of improvement and enhance their products and services, ultimately resulting in increased revenue (Kim et al., 2016). However, the large volume of reviews can make it challenging for both consumers and businesses to identify meaningful insights. Text mining techniques such as topic modeling and sentiment analysis can help to automate the analysis of large quantities of reviews and uncover patterns and insights more efficiently.

In this study, we conducted a review of recent studies on text mining performed on online consumer reviews (OCRs), with a focus on the machine learning techniques and their performances. We then applied and compared three most popular text mining approaches involving five models for OCRs. Our objective was to identify the most effective text mining methodologies to extract data analytics-based insights from large volumes of OCRs and to provide actionable recommendations for businesses and consumers.

## LITERATURE REVIEW

Online consumer reviews (OCRs) play a significant role in influencing the purchasing decisions of internet users, with 97-98% of internet users reporting that their purchases have been influenced by reviews (BrightLocal, 2023). The content and emotions expressed in these reviews are important for businesses to consider in order to improve their products and services. A study conducted on emotions in user reviews found that negative, objective reviews tend to have a greater influence on consumer decisions due to their perceived honesty (Lee et al., 2016). On the other hand, positive reviews have been found to increase bookings and revenue (Ye et al., 2009). The quantity of reviews can also have a positive impact on various aspects such as sales (Kim et al., 2016).

However, analyzing large datasets of OCRs, such as the dataset released by Yelp which comprises of over 8 million reviews, can be time-consuming process. Text mining techniques such as topic modeling and sentiment analysis have been used to address this challenge. Text mining is a rapidly growing field that involves the use of natural language processing and machine learning techniques to extract useful information from large amounts of unstructured text data. Text mining can be applied to the analysis of online consumer reviews, which can provide valuable insights into the sentiment and opinions of customers, therefore help businesses and organizations to identify patterns in customer opinions and sentiment and make data-driven decisions based on customer feedback.

Topic modeling or topic extraction is a text mining technique that is used to identify the main topics discussed in a large collection of documents. In addition, topic modeling can also assist in the representation of word vectors in deep learning models (Li et al., 2018). Topic modeling involves the use of algorithms such as Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) to identify patterns in the text data and group similar words together into topics. LDA is a popular method for extracting the most important topics to consumers (Huang et al., 2014; Zhang et al., 2022). LDA can be extended using word embedding algorithms to incorporate context into topic modeling, resulting in improved quality of extracted topics (Zhao et al., 2017; Yao et al., 2017; Ding, Dai, and Zhang, 2016; Moody, 2016). While LDA assumes a probabilistic generative process involving document-topic and topic-word distributions, NMF relies on matrix factorization, promoting sparsity and interpretability in the extracted topics. The choice between these methods depends on the specific characteristics of the data and the interpretability requirements of the analysis. LDA is often considered effective when dealing with large corpora and is widely used in the field of natural language processing. NMF is known for its ability to promote sparsity, which can result in more interpretable topics, especially in scenarios where the data has a natural non-negativity structure.

Sentiment analysis is another text mining technique that is commonly used to determine the overall sentiment or emotion conveyed in a piece of text. It involves the use of natural language processing and machine learning algorithms to classify text as positive, negative, or neutral based on the words and phrases used. This can be useful for identifying patterns in customer sentiment and understanding how customers feel about a particular product or service. Sentiment analysis of Twitter data was studied by (Medhat & Hassan, 2014) which shows how Twitter data can be used to extract useful insights. Popular Python libraries for sentiment analysis are TextBlob and VADER (Valence Aware Dictionary and sEntiment Reasoner). Studies (Kaur & Kaur, 2019; Jyoti & Bhardwaj, 2019) found that VADER performed better than TextBlob in sentiment analysis of Twitter data, especially when dealing with the informal language commonly used on the platform. However, the choice of algorithm depends on the context, specific use case and the data set being used.

Online consumer reviews from Yelp have been a popular and rich source of data for several studies that have used text mining techniques such as topic modeling and sentiment analysis. Zhang et al. (2016) used topic modeling to identify the main topics discussed in Yelp reviews of restaurants in New York City and found that the most common topics included food quality, service, and atmosphere. Liu et al. (2020) also found price and location of the restaurants were common topics of customer reviews in addition to food, service, atmosphere. Another study by (Hu et al., 2018) used sentiment analysis to classify Yelp reviews of hotels in Las Vegas as positive, negative, or neutral, and found that the majority of reviews were positive.

Topic classification is a supervised text mining technique that involves the process of categorizing text documents into predefined topic categories or classes based on the content of the text. It is used in various natural language processing (NLP) applications, such as text analysis, document organization, and recommendation systems. Convolutional Neural Networks (CNN) is an effective topic classification algorithm, particularly in text and image classification tasks (Rahul et al., 2018). CNN has also been used for sentiment analysis of reviews (Sharma et al., 2016). Other classifiers, such as Support Vector Machines (SVM) and Naive Bayes, have also been successful in sentiment analysis (Xu et al., 2014; Hemalatha & Ramathmika, 2019; Liu, 2020; Zhao, 2019).

The performance of machine learning models can be affected by various factors, including word count, dataset type, and corpus size (Zhao, 2019; Yang et al., 2018). In some cases, the combination of tools, such as spaCy or NLTK for data pre-processing, Condor for sentiment analysis, and SVM for text classification, can also improve performance (La Bella et al., 2018). The selection of a text pre-processing tool can have a significant impact on the efficiency of Natural Language Processing tasks. In this regard, the spaCy Natural Language Processing (NLP) library is a better alternative to NLTK (Natural Language ToolKit) for performing text pre-processing. (Omran & Treude, 2017).

In this study, we aim to verify the effectiveness of these proposed machine learning models for the text mining tasks of topic extraction, text classification, and sentiment analysis.

## RESEARCH METHODOLOGY

To identify the most effective text mining techniques for analyzing online customer reviews (OCRs), we applied a selection of text mining methods based on the literature review and applied them to a dataset comprising of Yelp restaurant reviews. By comparing and evaluating their performance, we aimed to identify the most effective approach for our specific analysis.

We applied and compared the following three methodologies involving five machine learning models in this study:

1. **Topic modeling using LDA vs. LDA2Vec.** Exploring topics in customer reviews helps businesses understand what customers are talking about and what their main concerns are. This information can be used to improve products or services, make better business decisions, and ultimately increase customer satisfaction and loyalty. This method involves exploring topics in customer reviews using an unsupervised learning model, Latent Dirichlet Allocation (LDA). We use two different approaches of LDA – LDA only and a hybrid of LDA and Word2vec, i.e. LDA2Vec.
2. **Text classification using CNN vs. CNN + GloVe.** The second method involves using a supervised learning method, Convolutional Neural Networks (CNN), to predict the labeled topics of the reviews. We use two different renditions of CNN - CNN only and CNN with GloVe, to see if word embedding technique like GloVe affects CNN's model performance.

3. **Sentiment analysis using CNN vs. SVM.** Analyzing sentiment in customer reviews helps businesses to understand customers' opinions and feelings towards products or services, which can inform decision-making for businesses. The third method compares two supervised learning models, CNN with SVM in classifying the sentiment of the reviews.

## Data Cleaning and Pre-Processing

In our study, we utilized the Yelp dataset challenge round 13 as the source of our data. This dataset contains a large collection of online restaurant reviews provided by Yelp, and it has been carefully curated and cleaned to ensure that it is suitable for use in various text mining tasks. With this dataset, we are able to access a wealth of information that can be used to understand consumer preferences and behaviors in the context of online restaurant reviews.

Data pre-processing for text mining involves a series of steps that are taken to prepare text data for use in text mining tasks. Some common steps that may be involved in data pre-processing for text mining include:

- Tokenization: This involves dividing the text into smaller units called tokens, which can be words, phrases, or other units of text.
- Stop word removal: This involves removing common words that are not useful for the text mining task, such as prepositions and articles.
- Stemming: This involves reducing words to their base form or root form, in order to eliminate variations caused by inflections and derivations. Stemming is simpler and faster than lemmatization. Sometimes stemming can result in words that are not real words or words that have lost their original meaning.
- Lemmatization: This involves reducing words to their base or root form (i.e. lemma), considering the context and part of speech of the word and uses a dictionary or knowledge of the language to convert the word to its canonical form. It is more accurate and sophisticated than stemming.
- N-gram: This involves text representation using a contiguous sequence of n items (words or characters) from a given sample of text or speech to predict the likelihood of certain words based on the statistical patterns and relationships between words.
- Part-of-speech tagging: This involves identifying the part of speech of each word in the text, such as nouns, verbs, and adjectives.
- Named entity recognition: This involves identifying and extracting named entities such as people, organizations, and locations from the text.

Tailoring our data pre-processing to the unique requirements of our text mining task and the characteristics of our dataset involves the following specific steps.

In our study, we began by parsing the reviews into 95,291 sentences. We then applied several data pre-processing steps to prepare the data for analysis. These steps included converting all text to lowercase, tokenizing the text, removing stop-words, white space, and punctuation, identifying and removing any errors, inconsistencies, or missing values in the data, and lemmatizing the words. We chose to use lemmatization rather than stemming because we were interested in preserving the meaning of the words for accuracy and correlation purposes. We also removed words that were less than 3 characters and more than 25 characters in length.

For the LDA and LDA2Vec topic modeling, we further filtered the data by removing short sentences that were less than 5 words in length. After the pre-processing, we were left with 1,777 sentences for further analysis. For the CNN and CNN + GloVe topic classification, our dataset initially consisted of 9,758 human-labeled sentences extracted from the original dataset. Each sentence was labeled with one of five topics: price, location, food, time, and service. The Price label contained 488 sentences, Location contained 1,054 sentences, Food contained 5,223 sentences, Time contained 333 sentences, and Service contained 2,660 sentences. Due to the uneven distribution of data, the dataset was unbalanced, and as a measure to overcome this issue, under-sampling was performed. As a result of this step, the dataset was balanced and consisted of 1,665 sentences for further analysis. We then applied the same data pre-processing steps as before and were left with 1,535 sentences for analysis.

For the CNN and SVM sentiment analysis, the dataset contained 9,762 labeled sentences. We had three labels: positive, neutral, and negative. Positive contained 5,590 sentences, neutral contained 1,297 sentences, and negative contained 2,875 sentences. This dataset was also unbalanced, so we balanced the data according to the label that contained the least number of sentences, which was the neutral label. The dataset was balanced to contain a total of 3,891 sentences, with 1,297 sentences for each label. For CNN, we performed the text pre-processing steps after the balancing and ended up with 3,810 sentences for further analysis. For SVM, we performed a similar pre-processing process with the addition of lemmatization using NLTK. After pre-processing, we were left with 2,895 sentences for further analysis.

**TABLE 1**
**DATA CLEANING AND PRE-PROCESSING**

| Model | Document Type | Before balancing | After balancing | Before Pre-processing | After Pre-processing |
|---|---|---|---|---|---|
| LDA and LDA2Vec for Topic Extraction | Sentence | NA | NA | 95291 | 1777 |
| CNN for Text Classification | Sentence | 9758 | 1665 | 1665 | 1535 |
| CNN + Glove for Text Classification | Sentence | 9758 | 1665 | 1665 | 1620 |
| CNN for Sentiment Classification | Sentence | 9762 | 3891 | 3891 | 3810 |
| SVM for Sentiment Classification | Sentence | 9762 | 3891 | 3891 | 2895 |

## Topic Extraction

*Latent Dirichlet Allocation (LDA)*

LDA is an unsupervised machine learning model used for the extraction of topics and grouping of related documents into these topics. When implementing LDA, we explored various parameter configurations such as the number of topics, the number of top keywords per topic, alpha and beta heuristics, hyperparameters, and the number of iterations. This iterative process allowed us to identify the optimal parameter settings that generated extracted results which were interpretable and meaningful.

*LDA + Word2Vec (i.e., LDA2Vec)*

When combined with word embedding technique, Word2vec, LDA can derive the semantic relationship between the words and documents through a map of word vectors by taking both the global and local context of the corpus into consideration. LDA2Vec showed promising results when compared to conventional LDA (Li et al., 2018; Moody, 2016). We want to determine whether the topics extracted by LDA2Vec are more cohesive than those extracted by conventional LDA. We trained a Word2vec model using skip-gram with 50-dimension word vectors. This was trained for 70 iterations. We set our topic number at 5 to extract five topics and trained the model for 400 epochs.

## Text Classification

*Convolutional Neural Networks (CNN)*

CNN is a supervised machine learning model commonly used for processing grid-like data including images, speech and text. It stands out for its performance in image recognition with an accuracy of up to 99.6% (Chauhan et al. 2018). CNN works as a deep learning model consisting multiple layers, including an input layer, convolutional and pooling layers that performs non-linear operations such as the Sigmoid and the Rectified Linear Unit (ReLU), and a fully connected output layer.

For our study we performed CNN in two different ways. The first used only CNN while the second included a GloVe word embedding layer. We used 70% for training and 30% for testing. We defined our vocabulary using the bag of words (BOW) method. The CNN model architecture included an embedding layer with a vocabulary size of 100, a convolutional layer with 1 filter and a kernel size of 3, and a rectified linear unit (ReLU) activation function. We also incorporated a MaxPooling1D layer with a pool size of 2, a flatten layer, and two dense layers with the ReLU and Sigmoid activation functions.

*CNN With GloVe*

CNN combined with word embedding methods like Word2vec (Zheng et al., 2016) and GloVe (Salinca, 2017; Zhong & Li, 2019) showed better results than conventional CNN. Word2vec is a neural network-based model that learns word embeddings by predicting the probability of a word given its context (continuous bag of words, CBOW) or predicting the context given a word (skip-gram). GloVe (Global Vectors) is another model that uses a co-occurrence matrix representing the frequency of each word's co-occurrence with every other word in a corpus to learn word embeddings. Word2vec tends to perform better on rare words because it can generate meaningful embeddings even for words with very few occurrences in the training corpus, while GloVe is better at capturing global semantic information, such as word associations and analogies.

We decided to implement CNN with GloVe and compare the result with CNN without GloVe. We used 70% for training and 30% for testing. We used the Keras TextVectorization library for vectorization. For the GloVe word embedding layer we found vectors for 1,464 words. The convolutional layer consisted of one dimension and 1 filter with a kernel_size of 3. The MaxPooling layer contains a pool size of 2, followed by a flatten layer. The Dense layer uses softmax as the activation. This model was trained for 400 epochs and with a batch size of 128.

### Sentiment Analysis

*Support Vector Machine (SVM)*

SVM has demonstrated success in various sentiment analysis tasks (Liu, 2020). Similar to CNN, SVM is a supervised machine learning model that operates on labeled data. In our study, SVM is applied to sentiment analysis, addressing a text classification challenge. While commonly employed for binary classification problems (Saifullah et al., 2021), SVM can also be adapted to multi-class classification models, as is the case in our study with three labels: positive, neutral, and negative. The dataset was partitioned into 70% for training and 30% for testing, and TF-IDF (Term Frequency-Inverse Document Frequency) was utilized for vectorization.

*Convolutional Neural Networks (CNN)*

We decided to employ the same CNN model used for text classification in sentiment analysis as well. The data split and parameters remained the same.

## RESULTS AND DISCUSSION

### Latent Dirichlet Allocation (LDA) for Topic Extraction

We found the following five topics and their top 10 weighted keywords using LDA:

- Topic 0: This topic seems to be related to Price with top 10 keywords to be "price, menu, portion, offer, size, free, cheap, excel, expens, expect."
- Topic 1: This topic seems to be related to Food with top 10 keywords to be "chicken, fri, sauc, chees, salad, tast, meat, rice, meat, sandwich"
- Topic 2: This topic seems to be related to Time with top 10 keywords to be "time, lunch, wait, night, minut, dinner, hour, call, check, busi"
- Topic 3: This topic seems to be related to Service with top 10 keywords to be "service, friendly, staff, nice, tabl, custom, clean, server, restaur, seat."

- Topic 4: This topic seems to be related to Location with top 10 keywords to be "restaur, locat, star, love, experi, park, look, street, local, spot."

**LDA + Word2Vec (LDA2Vec) for Topic Extraction**

We attempted to find the topics similar to LDA using LDA2Vec, however the results were not as coherent as LDA's. All the topics seem to contain many food-related words. We tried our best to get a general idea of the topics based on the frequent words or themes that are associated with each topic and summarize them in the following:

- Topic 0: This topic seems to be about food-related words such as "puppy", "cheesesteak", "cherry", "carb", "doughy", "cranberry", "concoction", and "asparagus". It also contains words that might relate to satisfaction, such as "satisfaction", "text rank", and "slide".
- Topic 1: This topic contains some food-related words such as "pumpkin", "guac", and "cheesesteak", but also includes words like "language practice", "character", "norm", and "gesture". It might relate to communication.
- Topic 2: This topic seems to contain words related to rewards, such as "voucher", "reward", "clientele", and "baseball". It also includes food-related words like "cherry", "carb", and "concoction".
- Topic 3: This topic contains a mix of food-related words like "cheesesteak", "cherry", and "sourdough" and non-food-related words like "image", "security", and "content". It might relate to food and security or food and technology.
- Topic 4: This topic seems to include food-related words such as "pot", "humus", "cheesesteak", and "lemongrass", and contains words that might relate to celebration or events, such as "trio", "presence", and "celebration".

There are several factors that could cause LDA2Vec in our study to perform poorer compared to LDA:

1. Dataset size: LDA2Vec may struggle with smaller datasets compared to LDA, as it relies on neural networks that require large amounts of data to learn effectively. Our dataset is a small one with only about 1,000 documents to train the model. Applying LDA2Vec on a larger dataset may generate better results.
2. Complexity: LDA2Vec is a more complex model than LDA, as it combines the LDA model with word embeddings and neural networks. This complexity can make it more difficult to train and optimize, and may require more computational resources. Our limited computing power was one of the factors that influenced the parameter choices we made. Cloud computing can increase computing power, scalability and flexibility which helps LDA2Vec to process large volumes of data more efficiently.
3. Hyperparameter tuning: LDA2Vec has several hyperparameters that need to be tuned, such as the number of topics, the size of the word embeddings, and the number of layers in the neural network. Finding the optimal values for these hyperparameters can be challenging and time-consuming.
4. The optimal number of topics: To determine the optimal number of topics, LDA2Vec can be trained with a range of topic number, then coherence scores can be calculated for each model to measure the interpretability, the optimal model can then be found at the point where the increase in coherence score begins to flatten out.

LDA2Vec is a promising approach for text mining that has the potential to outperform LDA in certain scenarios. However, it may not always be the best choice depending on the specific dataset and research question.

**CNN and CNN + GloVe for Topic Classification**

Our results reveal that all the classification models performed better than the baseline. For topic classification, CNN + GloVe performed better than CNN alone. This may be due to the added semantic information provided by the GloVe word embeddings. GloVe embeddings capture the semantic relationship between words, allowing the CNN model to better understand the meaning of the text, which can improve the accuracy of the classification task, especially when dealing with large datasets with complex and diverse text. Additionally, GloVe can help CNN overcome the problem of sparsity in the dataset by providing a continuous representation of words, which can help reduce overfitting and improve generalization of the model.

### CNN and SVM for Sentiment Analysis

On the other hand, for sentiment analysis, SVM outperformed CNN. SVM can work well with high-dimensional data, which is often the case in natural language processing (NLP) tasks, as text data can have a large number of features. In comparison, CNN may struggle with high-dimensional data and may require additional pre-processing techniques such as dimensionality reduction. Another reason for SVM's better performance could be the availability of labeled data. SVM typically requires less labeled data to train compared to CNN, and thus can perform better in cases where labeled data is scarce. Additionally, SVM is less prone to overfitting than CNN, which can lead to better generalization and performance on unseen data.

### Limitations

It is noteworthy that all models utilized small datasets due to the need for balancing and pre-processing, which may have impacted their performances. Further improvements may be achieved by optimizing the models' hyperparameters, using a larger dataset and more labeled data, and allocating more computing resources. Additionally, other popular models like K-nearest-neighbors (KNN) and XGBoost (XGB) may be explored for conducting topic modeling and sentiment analysis and comparing their performance with the models employed in this research.

## CONCLUSION

In this study we performed text mining to a Yelp dataset containing online consumer reviews, constructing various machine learning models to extract topics, classify text, and analyze sentiment. Utilizing these models for extracting insights can enhance recommender systems, furnish valuable information about sentiment and aspects crucial to consumers, aiding businesses in pinpointing areas for improvement or recognizing their strong suits. Consumers may also benefit from such automation, allowing for informed decisions amidst the deluge of reviews and improving their spending choices.

Our empirical study on the utilization of LDA2Vec for topic modeling suggests that the model's performance is subject to several factors, including dataset size, computational resources, and hyperparameter tuning. Concerning text classification, we observed that the combination of CNN with GloVe exhibited superior performance compared to CNN alone. In sentiment analysis, our findings suggested that SVM outperformed CNN.

In machine learning, the choice of algorithm often depends on the specific task and dataset characteristics. We strongly advocate for a systematic comparison of the performance of various algorithms on the same task to identify the most suitable one.

## REFERENCES


Al Omran, F.N.A., & Treude, C. (2017). Choosing an NLP Library for Analyzing Software Documentation: A Systematic Literature Review and a Series of Experiment. In *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)* (pp. 187–197). doi:10.1109/MSR.2017.42

BrightLocal. (2023, February 7). *Local Consumer Review Survey 2023. Insights / Research*. Retrieved from https://www.brightlocal.com/research/local-consumer-review-survey/

Chauhan, R., Ghanshala, K.K., & Joshi, R.C. (2018). Convolutional Neural Network (CNN) for Image Detection and Recognition. In *First International Conference on Secure Cyber Computing and Communication (ICSCCC)* (pp. 278–282). https://doi.org/10.1109/icsccc.2018.8703316

Ding, P., Dai, G., & Zhang, Y. (2016). Contextual-LDA: A Context Coherent Latent Topic Model For Mining Large Corpora. In *IEEE Second International Conference on Multimedia Big Data* (pp.420–425). doi: 10.1109/BigMM.2016.72.

Hemalatha, S., & Ramathmika, R. (2019). Sentiment analysis of yelp reviews by machine learning. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*. doi:10.1109/ICCS45141.2019.9065812

Hu, M., Lu, L., & Liu, B. (2018). Sentiment analysis of hotel reviews on Yelp. *Journal of Hospitality and Tourism Technology*, *9*(2), 131–148.

Huang, J., Rogers, S., & Joo, E. (2014). Improving Restaurants by Extracting Subtopics from Yelp Reviews. In *Social Media Expo* (pp. 1–5).

Jyoti, R., & Bhardwaj, P. (2019). Sentiment analysis of Twitter data using TextBlob and VADER: A comparative study. *International Journal of Engineering and Advanced Technology*, *8*(3), 652–656.

Kaur, J., & Kaur, M. (2019). A comparative study of TextBlob and VADER for sentiment analysis of Twitter data. *International Journal of Engineering and Advanced Technology*, *8*(3), 646–651.

Kim, W.G., Li, J., & Brymer, R.A. (2019). The impact of social media reviews on restaurant performance: The moderating role of excellence certificate. *International Journal of Hospitality Management*, *55*, 41–51. doi: 10.1016/j.ijhm.2015.10.004

La Bella, A., Colladon, A.F., Battistoni, E., Castellan, S., & Francucci, M. (2018). Assessing Perceived Organizational Leadership Styles Through Twitter Text Mining. *Journal of the Association for Information Science and Technology*, pp. 1–31.

Lee, M., Jeong, M., & Lee, J. (2017). Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website: A text mining approach. *International Journal of Contemporary Hospitality Management*, pp. 762–783.

Li, Q., Li, S., Hu, J., Zhang, S., & Hu, J. (2018). Tourism Review Sentiment Classification Using a Bidirectional Recurrent Neural Network with an Attention Mechanism and Topic-Enriched Word Vectors. *Sustainability*, pp. 1–15. MDPI. doi: 10.3390/su10093313

Liu, S. (2020). *Sentiment Analysis of Yelp Reviews: A Comparison of Techniques and Models*. arXiv preprint arXiv:2004.13851.

Medhat, W., & Hassan, A.E. (2014). *Sentiment analysis of Twitter data*. arXiv preprint arXiv:1409.2132.

Moody, C. (2016, May 27). Introducing our hybrid lda2vec algorithm. [Blog post]. *Multithreaded*. Retrieved from https://multithreaded.stitchfix.com/blog/2016/05/27/lda2vec/

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135.

Saifullah, S., Fauziyah, Y., & Aribowo, A.S. (2021). Comparison of machine learning for sentiment analysis in detecting anxiety based on social media data. *Journal Informatic*, *15*(1), 1–10. https://doi.org/10.26555/jifo.v15i1.a20111

Salinca, A. (2017). Convolutional Neural Networks for Sentiment Classification on Business Reviews. *Proceedings of IJCAI Workshop on Semantic Machine Learning*. https://doi.org/10.48550/arXiv.1710.05978

Sharma, R., Tripathi, S., Sahu, S., Mittal, S., & Anand, A. (2016). A novel approach for sentiment analysis using deep learning techniques. *Procedia Computer Science*, *78*, 381–386. https://doi.org/10.1016/j.procs.2016.02.079

Xu, Y., Wu, X., & Wang, Q. (n.d.). *Sentiment Analysis of Yelp's Ratings Based on Text Reviews* (pp. 1–5).

Yang, X., Macdonald, C., & Ounis, I. (2016, July 21). Using word embeddings in Twitter election classification. *Proceedings of Neu-IR'16: SIGIR Workshop on Neural Information Retrieval*, Pisa, Italy. https://doi.org/10.48550/arXiv.1606.07006

Yao, L., Zhang, Y., Chen, Q., Qian, H., Wei, B., & Hu, Z. (2017). *Mining coherent topics in documents using word embeddings and large-scale text data*. Elsevier Ltd. https://doi.org/10.1016/j.procs.2017.01.071

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, *28*(1), 180–182. https://doi.org/10.1016/j.ijhm.2008.06.011

Zhang, S., Ly, L., Mach, N., & Amaya, C. (2022). Topic modeling and sentiment analysis of Yelp restaurant reviews. *International Journal of Information Systems in the Service Sector (IJISSS)*, *14*(1).

Zhang, Y., Li, X., & Sun, A. (2016). A topic model for Yelp restaurant reviews. In *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 4373–4379).

Zhao, S. (2019). *Thumb Up or Down? A Text-Mining Approach of Understanding Consumers through Reviews* (pp. 1–21). Decision Sciences Institute.

Zhao, S., Han, S., Meng, R., He, D., & Zhang, D. (2017). Learning Semantic Representation from Restaurant Reviews: A Study of Yelp Dataset. *iConference 2017 Proceedings, 2*, 159–162. https://doi.org/10.9776/17367

Zheng, X., Li, Y., & Li, L. (2016). Convolutional neural networks with word embeddings for sentiment analysis of Yelp reviews. In *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 146–153). IEEE.

Zhong, J., & Li, W. (2019). *Predicting Customer Call Intent by Analyzing Phone Call Transcripts Based on CNN for Multi-Class Classification* (pp. 9–20). CS & IT-CSCP.