A Hybrid Approach of Opinion Mining and Comparative Linguistic Analysis of Restaurant Reviews

Salim Sazzed
Old Dominion University
Norfolk, VA, USA
ssazz001@odu.edu

Abstract

The existing research on sentiment analysis mainly utilized data curated in limited geographical regions and demography (e.g., USA, UK, China) due to commercial interest and availability of review data. Since the user's attitudes and preferences can be affected by numerous sociocultural factors and demographic characteristics, it is necessary to have annotated review datasets belong to various demography. In this work, we first construct a review dataset *BanglaRestaurant* that contains over 2300 customer reviews towards a number of Bangladeshi restaurants. Then, we present a hybrid methodology that yields improvement over the best performing lexicon-based and machine learning (ML) based classifier without using any labeled data. Finally, we investigate how the demography (i.e., geography and nativeness in English) of users affect the linguistic characteristics of the reviews by contrasting two datasets, *BanglaRestaurant* and *Yelp*. The comparative results demonstrate the efficacy of the proposed hybrid approach. The data analysis reveals that demography plays an influential role in the linguistic aspects of reviews.

1 Introduction

Sentiment analysis or opinion mining refers to the process of identifying opinions or sentiments expressed (e.g., positive, negative) in a text document (Liu, 2012) . The lexicon-based method and machine learning (ML) based method are the two dominant approaches for opinion mining; although, their combinations have been also explored by the researchers. To evaluate the polarity of a piece of text, the lexicon-based methods rely on the sentiment lexicon comprised of opinion-conveying positive or negative terms and a set of rules (Turney, 2002; Sazzed, 2020b). For the lexicon-based methods, the laborious steps of data labeling are not required. The supervised machine learning (ML) based approaches derive the relationship between features of the text segments and the opinions expressed in the writing in a supervised fashion (Pang et al., 2002; Sazzed and Jayarathna, 2019). Therefore, supervised ML classifiers require a significant amount of annotated data for training a predictive model for determining the polarity of a text document. Labeling a large amount of text data is not only a challenging but also a tedious and costly process (Sazzed and Jayarathna, 2021; Sazzed, 2020a). Researchers also investigated the combinations of both lexicon-based and ML-based approaches to form a hybrid method (Kolchyna et al., 2015; Mendon et al., 2021; Sazzed, 2021).

Mining customer opinions towards restaurants has attained popularity in recent years due to its impact on business growth and sustainability. Researchers performed a number of studies using the *Yelp* restaurant review and several other datasets. However, existing research on opinion mining in review datasets mainly focused on data curated in some specific demography (e.g., developed countries or countries with large consumer markets) due to commercial interest and abundance of user-generated review data. As shown in (Nakayama and Wan, 2019), the trait and preferences of users can vary across demography. Therefore, it is important to generate annotated content for various demography (e.g., geography, user, or language) and analyze them to identify the differences, which can ultimately help better decision-making. Nowadays, with the increasing accessibility of the Internet and the popularity of social media, opinion data are increasingly becoming available in many other geographies (e.g., Bangladesh) and languages.

Therefore, the main objective of this work is to create a restaurant review dataset from less-explored demography and introduce a new methodology to improve the performance of sentiment

Proceedings of Recent Advances in Natural Language Processing, pages 1281–1288
Sep 1–3, 2021.
https://doi.org/10.26615/978-954-452-072-4_144

classification. Besides, this study aims to explore how the demography affects the linguistic characteristics of reviews.

We create a restaurant review dataset, *BanglaRestaurant*, which contains more than 2300 customer reviews toward various Bangladeshi restaurants. We employ both the lexicon-based and ML-based methods to classify customer's sentiments in the *BanglaRestaurant* dataset. To improve the performance of sentiment classification, a hybrid methodology is introduced that leverages a lexicon-based method and an ML-based classifier. We observe an improvement of the F1 score by employing the proposed hybrid approach.

We investigate the characteristics of reviews belong to the *BanglaRestaurant* dataset written by non-native English speakers and *Yelp* reviews written by English native speakers. The comparative analysis reveals that demography (i.e., nativeness of language and geography) has influences on the various linguistic features of reviews.

1.1 Contributions

The contributions of this paper can be summarized as follows:

- We create a Bangladeshi restaurant dataset consists of over 2300 customer reviews [1].
- We propose a hybrid approach that improves the performance of sentiment classification by combining the lexicon-based method and supervised ML classifier.
- We analyze the characteristics of two restaurant review datasets curated in different demography.

2 Sentiment Analysis in Restaurant Review Datasets

Kang et al. (2012) created a sentiment lexicon and proposed an improved Naive Bayes (NB) based method for sentiment analysis in a restaurant dataset. Blair-Goldensohn et al. (2008) introduced a sentiment summarizer system where a summary is built by extracting relevant aspects of a service, aggregating the sentiment per aspect, and selecting aspect-relevant text. An attention-based Long Short-Term Memory (LSTM) network was proposed in (Wang et al., 2016) for aspect-level sentiment classification. Gan et al. (2017) analyzed how various attributes influence customers sentiments on restaurant star ratings. Zhang et al. (2011) incorporated ML-based techniques such as NB and SVM to automatically classify user reviews as positive or negative from online Cantonese-written restaurant reviews.

Zahoor et al. (2020) created a restaurant dataset by collecting over 4000 customer reviews of various restaurants located in Pakistan. Sasmita et al. (2017) performed aspect-based sentiment analysis (ABSA) in Indonesian restaurant reviews. They performed both the (i) aspect extraction and (ii) aspect sentiment orientation classification.

Xue et al. (2017) identified aspect categories and extracted aspect-terms from the user-generated reviews. The authors proposed a multi-task learning model based on neural networks and observed improved performance over the models trained separately on three public datasets. Ahiladas et al. (2015) utilized named entity recognition (NER) and typed dependency techniques to identify different types of food and the opinions associated with them. Tian et al. (2021) performed a case study on *Yelp* restaurant review data to find what affects restaurant customer's sentiments. Besides, they noticed consumers rate restaurant service more often than the food quality. Jia (2018) proposed an integrated approach that leverages text mining and empirical modeling to correlate ratings with reviews. The author examined 49,080 pairs of restaurant ratings and reviews from Dianping.com (a Chinese online review community) to identify high-frequency words, major topics, and subtopics. Xiang et al. (2019) presented an LSTM based architecture LSTM-SAT for sentiment analysis of Cantonese style text by incorporating sentiment knowledge into the attention mechanism in the LSTM.

3 BanglaRestaurant Dataset

3.1 Data Collection

The restaurant reviews are manually collected from the restaurant's Facebook pages. We find the reviews are written in English, Bengali (i.e., the native language of Bangladesh), and Romanized-Bengali (Bengali text in Latin alphabet). English is the main foreign language in Bangladesh which is taught in schools and colleges. Besides, English is frequently used in government administration, educational institutions, courts, businesses, and media of the country. People often use English for

[1] `https://github.com/sazzadcsedu/BanglaRestaurant.git`

expressing their opinions and feelings on social media as it is more convenient to write English text than Bengali. For example, Bengali has 50 letters (11 vowels and 39 consonants) compared to 26 letters in English. Since this study focuses on the reviews written in English, the final dataset contains only the English reviews.

3.2 Data Annotation

We annotate the reviews based on the reviewer's recommendations; if the reviewer recommends the restaurant, then the corresponding review belongs to the positive class; Otherwise, it goes to the negative class. The final restaurant dataset contains a total of 2315 reviews, 1702 positive reviews (i.e., recommended by customer), and 613 negative reviews (not recommended).

4 Proposed Methodology

To determine the semantic orientations of the reviews, we employ both the lexicon-based and ML-based methods, as each of them has certain advantages over the other.

4.1 Lexicon-Based Approaches

We employ four lexicon-based methods: VADER (Hutto and Gilbert, 2014), TextBlob, LRSentiA (Sazzed and Jayarathna, 2021) and SentiStrength (Thelwall et al., 2010) for classifying sentiment from unlabeled data. A non-negative polarity score by these methods is considered as a positive prediction (except VADER, where the compound score is used).

4.1.1 VADER

VADER (Valence Aware Dictionary for Sentiment Reasoning) is a lexicon and rule-based sentiment analysis tool specifically attuned to determine sentiments expressed in social media. In VADER, a compound score indicates the semantic orientation of a review. For binary classification, a non-negative compound score refers to a positive prediction.

4.1.2 LRSentiA

LRSentiA is a lexicon and rule-based method that can classify opinions expressed in unlabeled data. LRSentiA utilizes a binary-level opinion lexicon (Liu, 2010) and a set of linguistic rules to determine the polarity of a review [2].

[2] https://github.com/sazzadcsedu/SSentiA.git

4.1.3 TextBlob

TextBlob is a Python library for processing textual data. The predicted polarity score of a review is within the range of [-1, +1], where -1 indicates strongly negative, and +1 means strongly positive.

4.1.4 SentiStrength

SentiStrength predicts the strength of positive and negative sentiments in short texts. The range of sentiment value of negative sentiment could be between -1 to -5; For the positive sentiment, the score can range between +1 and +5.

4.2 Machine Learning Approaches

4.2.1 ML Classifiers

In this work, we employ five popular supervised ML classifiers: Logistic Regression (LR), Ridge Regression (RR), Support Vector Machine (SVM), Random Forest (RF), and Extra Tree Classifier (ET) for identifying the polarity of the customer review.

4.2.2 Experimental Settings

The review texts are segmented into words and converted to a matrix of term frequency-inverse document frequency (TF-IDF) features. We calculate the TF-IDF score for all words in the review using the scikit-learn library (Pedregosa et al., 2011), and the resultant matrix is feed to supervised ML classifiers. To evaluate the performance of various ML classifiers, we use 10-fold cross-validation. For all the ML classifiers, the default parameter settings of the scikit-learn library (Pedregosa et al., 2011) with class-balanced weights are used.

4.3 Hybrid Approach

We present a hybrid methodology for sentiment classification by leveraging both the lexicon-based and ML-based approaches. Based on the predicted polarity scores of the reviews determined by the lexicon-based method LRSentiA (Sazzed and Jayarathna, 2021), we categorize them into three groups.

1. *Minimal opinion group* (MOG): When LRSentiA assigns a polarity score of 0 to a review, it falls into the MOG category. A review with a 0 polarity score is considered as a positive prediction assuming that a negative review has a higher chance of having negative polarity scores than a positive review with a positive polarity score. Thus, predictions with non-negative polarity scores are considered positive predictions.

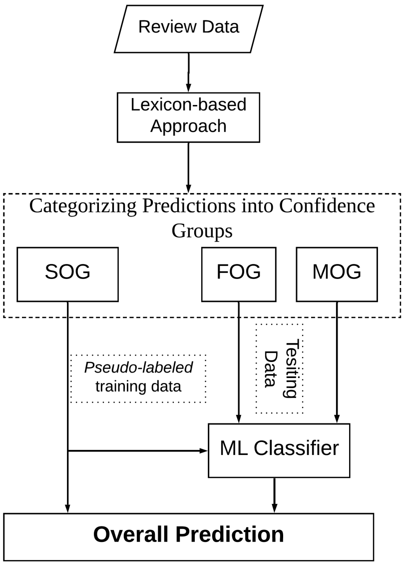

Figure 1: The proposed hybrid approach

2. *Fair opinion group* (FOG): When the predicted polarity score of a review is between $< -2, +2 >$, it belongs to the FOG group.

3. *Strong opinion group* (SOG): The reviews with polarity scores above +2 or less than -2 fall into SOG.

We assume if a lexicon-based method predicts the class of a review with a high polarity score (i.e., highly positive (> 2) or highly negative score (< -2)), it is highly probable that prediction is correct. As the lexicon-based method relies on the polarity of individual opinion words, if the overall polarity score is strongly positive or negative, then the review consists of mostly positive aspects (high positive score) or negative aspects (high negative score); thus, the prediction is probably correct.

After excluding the reviews belong to *strong opinion group* (SOG), the remaining two groups contain reviews which the lexicon-based method cannot distinguish confidently based on the opinion words present in the review. This scenario could happen due to various reasons, such as lexicon coverage problems or the complexity of natural languages. These groups are more prone to misclassification by the lexicon-based methods.

ML classifiers have been successfully applied in numerous problems in varying domains when data is noisy or explicit rules can not separate the classes well. Since supervised ML classifiers are capable of characterizing the best mapping from input to output, we employ an ML classifier for the reviews that require learning the implicit pattern from the labeled data rather than just using the polarity of the individual opinion words to find overall sentiment.

Reviews	True Class
This time the food was not good and food was cold and not tasty....	Negative
minimum Cost maximum fun. nice place, good food and behaviour is also good....	Positive
It is a very nice place for eating. Last night we arranged our daughter's birthday party. It was excellent. All guests are happy with their food & services. It was really good & more than our expectation. We are fully satisfied with their food & services. Thanks to cafe Rio. We really loved it. Keep it up...	Positive
This is by far worst buffet in Dhaka. There are various budget buffet with common and boring items like flavors, premium club, buffet king, Ratatouille etc. They at least serve better foods. But cafe rio, a big disappointment. They banned me and other customers who comments against their restaurant. They are shame in the sector of food in Bangladesh. Yack	Negative
Well decorated,tasty foods, good service.Highly recommended.	Positive

Figure 2: Sample reviews from *BanglaRestaurant* dataset

The overall predictions by the hybrid method consist of predictions by LRSentiA for the highly polar reviews and an ML classifier for the remaining reviews.

5 Impact of Demography in Linguistic Attributes of Review

We analyze whether the demography of the reviews has any impact on the linguistic characteristics of the reviews. We consider two datasets, which differ in the following demographic aspects, geography (i.e., the location where reviews were written) and language nativeness of speakers (i.e., whether native or non-native speakers of English wrote reviews). Our developed *BanglaRestaurant* corpus is curated in Bangladesh, where people are non-native speakers of English (sample reviews shown in Figure 2). We contrast this dataset with the *Yelp* restaurant reviews, which were written by USA-based (mostly) English native speakers (sample reviews shown in Figure 3).

The *Yelp* dataset contains 6860 positive comments and 1676 negative comments in contrast to 1702 positive samples and 613 negative samples in *BanglaRestaurant* dataset. To avoid any kind of influence of class distribution and dataset size, we use the same number of reviews from both the *BanglaRestaurant* and *Yelp* datasets. For the *Yelp* dataset, we randomly shuffle and then select 1702 positive samples and 613 negative reviews out of 6860 positive and 1676 negative reviews present.

Reviews	True Class
I am a huge fan of Chipotle but I will never come back to this location. The staff looks miserable and are SLOW. The cashier decided she would rather clean tables then ring people up so we waited for her to come back. Most of the tables were already clean so she could of waited to clean. The rice was crunchy like it wasn't cooked all the way and the steak was really chewy. Never again!	Negative
I love scallops, and the scallops at The Breadfruit were simply the best. On the weekend, The Breadfruit is open till 11pm and its Rum Bar is open till midnight, which makes this a great late-night haunt: pick your favorite rum drink and share a scallop appetizer for the perfect lite end to the evening. The servers were excellent and handled my gluten allergy easily, offering suggestions and slight modifications to the dishes. The Breadfruit is located at 108 East Pierce, which is south of Roosevelt and just east of First Street. Put this place as a must on your late-night hit list.	Positive

Figure 3: Sample reviews from the *Yelp* restaurant dataset

We analyze the following characteristics of the reviews in two datasets. For the *Yelp* dataset, the numbers represent the average results of five random selections of 1702 positive comments and 613 negative comments.

1. Word count of the corpus: The numbers of words present in both corpora are provided.

2. Sentence count in the corpus: We report the number of sentences present in both datasets.

3. Average review length (word-level): The average review length in word-level indicates the average number of words present in a review of a corpus.

4. Average review length (sentence-level): The average review length in sentence-level refers to the average number of words present in a review of a corpus.

5. Average sentence length: This metric provides the average number of words each sentence contains in reviews that belong to a corpus.

6. Coverage of a lexicon: Furthermore, we compute the lexicon coverage of two English lexicons (Liu, 2010) and (Hutto and Gilbert, 2014) in both datasets. The lexicon coverage can assess the presence of diverse opinion words in the reviews, which indicates the vocabulary range of users.

7. Usage of the complex sentence in reviews: Besides, we study the complexity of the reviews at the sentence level. A complex sentence usually contains one or more dependent (subordinate) clauses and one or more independent clauses. A subordinating conjunction is a word or phrase that connects a dependent clause to an independent clause. Some examples of subordinating conjunctions are, *although, as, because, before, how, if, once, since, etc.*. We examine the presence of 50 common subordinating conjunctions [3] in both corpora to analyze the complexity of the reviews.

6 Results

Method	Pre.	Rec.	F1	Acc.
SentiStrength	0.896	0.789	0.839	88.0%
TextBlob	0.896	0.821	0.857	88.7%
VADER	0.895	0.824	0.858	89.2%
LRSentiA	0.901	0.822	0.860	89.5%
ET	0.846	0.832	0.834	87.9%
RF	0.855	0.833	0.840	88.0%
LR	0.878	0.904	0.891	91.4%
RR	0.884	0.901	0.893	91.7%
SVM	0.882	0.903	0.893	91.5%

Table 1: Performances of Various Lexicon-based and ML-based Methods in *BanglaRestaurant* Dataset

Group	Pre.	Rec.	F1	Acc.
MOG	NA	NA	NA	61.65% (201/326)
FOG	0.86	0.84	0.85	87.04% (578/664)
SOG	0.97	0.95	0.96	97.58% (1293/1325)
Overall	0.90	0.82	0.86	89.5% (2072/2315)

Table 2: Performance of The Best Lexicon-based Method LRSentiA in *BanglaRestaurant* Dataset

Table 1 reveals that TextBlob, VADER, and LRSentiA perform similarly, where the SentiStrength yields comparatively lower F1 score and accuracy. SVM, LR, and RR provide a similar F1 score of around 0.89 and an accuracy of 91%, which is a bit higher than the top lexicon-based method, LRSentiA. Decision tree-based ML classifiers such as DT and ET provide comparatively low accuracy and F1 score compared to other ML classifiers.

[3] https://github.com/sazzadcsedu/50SubordinateConjunctions.git

Features	***BanglaRestaurant***	***Yelp***
Total number of words in corpus	61523	295781.2
Total number of sentences in corpus	7258	22588.6
Avg. number of words/review	26.575	127.767
Avg. number of sentences/review	3.1352	9.757
Avg. number of words/sentence	8.47	13.09
Number of opinion words (Hu-Liu)	4377	16112.4
Lexicon coverage (Hu-Liu)	7.1%	5.4%
Number of opinion words (VADER)	4655	16429.6
Lexicon coverage (VADER)	7.56%	5.55%
Subordinating conjunctions in corpus	1214(61523)	8704.4(295781.2)
Subordinating conjunctions per reviews	0.52	3.76

Table 3: The Various Linguistics Attributes of Reviews Belong to *Yelp* and *BanglaRestaurant* Datasets

Method	**F1 Score**	**Accuracy**
LRSentiA	0.755	79.6%
ET	0.744	79.6%
RF	0.776	78.8%
RR	0.831	85.1%
LR	0.840	85.7%
SVM	**0.845**	**86.3%**

Table 4: Accuracy and F1 Scores of Various ML Classifiers in MOG and FOG Groups (990 Reviews)

Method	**F1 Score**	**Accuracy**
Hybrid-LR	0.902	92.5%
Hybrid-SVM	**0.915**	**92.75%**

Table 5: Accuracy and F1 Scores of The Hybrid Approach Integrating Two Best Performing ML Classifiers

Table 2 exhibits that lexicon-based method LRSentiA fails to classify the reviews correctly in many cases when the predicted polarity score is low (i.e., between -2 and +2, inclusive). In the MOG and FOG groups, only 779 out of 990 reviews are classified correctly. We notice when the lexicon-based method predicts with high polarity score, it is accurate in most cases. Among 1325 reviews, the predictions of LRSentiA are true for 1293 cases with an accuracy of 97.5%.

Table 4 presents the performance of ML classifiers in the complex subsets (MOG and FOG) of *BanglaRestaurant* reviews (i.e., 990 reviews out of 2315), which the lexicon-based method fails to not discern correctly. We find that most of the ML classifiers yield better performance compared to the lexicon-based classifier. The best performing classifier SVM increases F1 score 12% over the lexicon-based method in the MOG and FOG confidence groups.

Table 5 shows that the proposed hybrid method enhances the F1 score and accuracy of classification by integrating an ML classifier such as LR or SVM with the lexicon-based method. While the best lexicon-based and ML-based methods show F1 scores of 0.86 and 0.893, respectively, the hybrid approach incorporating the SVM classifier attains an F1 score of 0.91.

From Table 3, we observe that *Yelp* review lengths are much higher in both word and sentence levels. The sentiment lexicon shows higher coverage in *BanglaRestaurant*; The presence of the subordinate clause, which refers to the complex sentence, is more obvious in the *Yelp* dataset.

7 Discussion

We observe in the *BanglaRestaurant* dataset performance of the lexicon-based approach is close to the ML-based classifiers. The best macro F1 score is obtained from the SVM classifier, which is around 89%; The most accurate lexicon-based method LRSentiA achieves an accuracy of 86%. This result is expected as supervised ML classifiers usually perform better than lexicon-based methods.

From Table 2, it is evident that the lexicon-based method is very effective when the review polarity is easily distinguishable either as positive or negative. If a user review is comprised of mixed opinions towards various entities or sentiment is not obvious, it is often difficult to assign the overall polarity using the lexicon-based method. In contrast, the ML classifiers learn implicit patterns from training data, thus, are capable of determining the overall sentiment of a review even though the opinion is not apparent. Thus combining both the lexicon and ML-based classifiers in the sentiment classification framework improves the performance.

We notice the hybrid approach yields an overall F1 score of 0.915 in the *BanglaRestraurant* corpus, an improvement of 6% (0.860) over the best lexicon-based method and 3% (0.892) over the best ML-based classifier. Although the increase of the F1 score is not much compared to the best ML classifier, the main advantage of the proposed hybrid approach is that it does not require any annotated data. Review data is usually readily available on the web; the primary challenge is to label the data. Thus, the proposed hybrid approach can be very effective for addressing the data annotation difficulties. We find the features characteristics of the reviews are distinct in *BanglaRestaurant* and *Yelp* datasets, which represent data from different demography.

8 Summary and Conclusions

In this work, we introduce a hybrid approach for sentiment classification in a newly created *BanglaRestaurant* dataset. The proposed hybrid approach combines the lexicon-based method LRSentiA with the SVM classifier to improve the performance of sentiment classification. The results suggest that lexicon-based methods are mainly effective at classifying reviews that contain strong opinions. However, they struggle to determine sentiment when the polarity is not obvious. Hence, it is necessary to incorporate an ML classifier that is robust for complex cases.

In addition, we provide a comparative analysis of review data curated in different demography. We investigate various linguistic features of reviews that belong to these two datasets. The first dataset contains (i.e., *BanglaRestaurant*) reviews written by non-native English speakers of Bangladesh; while, reviews of the other dataset (i.e.,*Yelp*) were written by (mostly) English native speakers located in the USA. We observe differences in the various linguistic characteristics of reviews in these two datasets.

Acknowledgments

The author likes to thank Md. Samiul Basir Tasin and MD Shafin Islam Rudro for collecting *BanglaRestaurant* review data. The conference registration fee was supported by the ISAB VISA Scholarship of Old Dominion University.

References

Burusothman Ahiladas, Paraneetharan Saravanaperumal, Sanjith Balachandran, Thamayanthy Sripalan, and Surangika Ranathunga. 2015. Ruchi: Rating individual food items in restaurant reviews. In *Proceedings of the 12th International Conference on Natural Language Processing*, pages 209–214.

Sasha Blair-Goldensohn, Kerry Hannan, Ryan T. McDonald, Tyler Neylon, George A. Reis, and Jeff Reynar. 2008. Building a sentiment summarizer for local service reviews.

Qiwei Gan, Bo H. Ferns, Yang Yu, and Lei Jin. 2017. A text mining and multidimensional sentiment analysis of online restaurant reviews. *Journal of Quality Assurance in Hospitality & Tourism*, 18(4):465–492.

Clayton J. Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *ICWSM*.

Susan Jia. 2018. Behind the ratings: Text mining of restaurant customers' online reviews. *International Journal of Market Research*, 60(6):561–572.

Hanhoon Kang, Seong Joon Yoo, and Dongil Han. 2012. Senti-lexicon and improved naïve bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, 39(5):6000 – 6010.

Olga Kolchyna, Tharsis TP Souza, Philip Treleaven, and Tomaso Aste. 2015. Twitter sentiment analysis: Lexicon method, machine learning method and their combination. *arXiv preprint arXiv:1507.00955*.

Bing Liu. 2010. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing, Second Edition. Taylor and Francis Group, Boca*.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Shalak Mendon, Pankaj Dutta, Abhishek Behl, and Stefan Lessmann. 2021. A hybrid approach of machine learning and lexicons to sentiment analysis: enhanced insights from twitter data of natural disasters. *Information Systems Frontiers*, pages 1–24.

Makoto Nakayama and Yun Wan. 2019. The cultural impact on social commerce: A sentiment analysis on yelp ethnic restaurant reviews. *Information & Management*, 56(2):271–279.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Dhanang Hadhi Sasmita, Alfan F Wicaksono, Samuel Louvan, and Mirna Adriani. 2017. Unsupervised aspect-based sentiment analysis on indonesian restaurant reviews. In *2017 International Conference on Asian Language Processing (IALP)*, pages 383–386. IEEE.

Salim Sazzed. 2020a. Cross-lingual sentiment classification in low-resource bengali language. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pages 50–60.

Salim Sazzed. 2020b. Development of sentiment lexicon in bengali utilizing corpus and cross-lingual resources. In *2020 IEEE 21st International conference on information reuse and integration for data science (IRI)*, pages 237–244. IEEE.

Salim Sazzed. 2021. Improving sentiment classification in low-resource bengali language utilizing cross-lingual self-supervised learning. In *International Conference on Applications of Natural Language to Information Systems*, pages 218–230. Springer.

Salim Sazzed and Sampath Jayarathna. 2019. A sentiment classification in bengali and machine translated english corpus. In *2019 IEEE 20th international conference on information reuse and integration for data science (IRI)*, pages 107–114. IEEE.

Salim Sazzed and Sampath Jayarathna. 2021. Ssentia: A self-supervised sentiment analyzer for classification from unlabeled data. *Machine Learning with Applications*, page 100026.

Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, 61(12):2544–2558.

Guang Tian, Liang Lu, and Christopher McIntosh. 2021. What factors affect consumers' dining sentiments and their ratings: Evidence from restaurant online review data. *Food Quality and Preference*, 88:104060.

Peter D. Turney. 2002. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424, Stroudsburg, PA, USA. Association for Computational Linguistics.

Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. Attention-based lstm for aspect-level sentiment classification. In *EMNLP*.

Rong Xiang, Ying Jiao, and Qin Lu. 2019. Sentiment augmented attention network for cantonese restaurant review analysis. In *Proceedings of WISDOM'19: Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM'19)*.

Wei Xue, Wubai Zhou, Tao Li, and Qing Wang. 2017. Mtna: a neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 151–156.

Kanwal Zahoor, Narmeen Zakaria Bawany, and Soomaiya Hamid. 2020. Sentiment analysis and classification of restaurant reviews using machine learning. In *2020 21st International Arab Conference on Information Technology (ACIT)*, pages 1–6. IEEE.

Ziqiong Zhang, Qiang Ye, Zili Zhang, and Yijun Li. 2011. Sentiment classification of internet restaurant reviews written in cantonese. *Expert Systems with Applications*, 38(6):7674 – 7682.