International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 03 | Mar 2024 www.irjet.net p-ISSN: 2395-0072

THE MACHINE LEARNING-BASED ANALYSIS OF SENTIMENT IDENTIFICATION FOR POSTS

Savita Tripathi[1], Mr. Sambhav Agarwal[2]

[1]M.Tech, Computer Science and Engineering, SR Institute of Management & Technology, Lucknow, India
[2]Assistant Professor, Computer Science and Engineering, SR Institute of Management & Technology, Lucknow

---

Abstract - *Sentiment analysis is a field of research that studies people's opinions about different things, such as products, social and political events, and problems. This type of analysis has become increasingly popular because it can help stakeholders make better decisions based on public opinion. Opinion mining is one way to gather information from sources like search engines, web blogs, Twitter, and social networks. However, because there are so many tweets available online in unstructured text form, it can be difficult to analyze them manually. To solve this problem, researchers use computational strategies that involve identifying sentiment-bearing words in the text. There are many different methods for doing this using machine-learning techniques like Bag-of-Words (BoW) representation. In this study specifically, the researchers used a lexicon-based approach to automatically identify sentiment in tweets collected from Twitter's public domain. They also applied three different machine learning algorithms – Naive Bayes (NB), Maximum Entropy (ME), and Support Vector Machines (SVM) – to see which was most effective at classifying the tweets by sentiment. The experiments showed that both NB with Laplace smoothing and SVM were effective classifiers when using certain features like unigrams or Part-of-Speech (POS). Overall, sentiment analysis is an important tool for understanding public opinion on various topics through user-generated content on platforms like Twitter.*

Key Words: Bag-of-Words (BoW), Lexicon, Machine Learning Algorithms, Laplace Smoothing, Part-of-Speech (POS).

1. INTRODUCTION

The process of utilizing Twitter for automatic sentiment identification involves a series of vital steps. The first step is to gather a large dataset of tweets using Twitter's API, with a focus on specific keywords, hashtags, or timelines of interest. Once this data has been collected, it undergoes preprocessing, which involves removing noise such as URLs, special characters, and stopwords. Additionally, the text is tokenized and normalized. After preprocessing, feature extraction takes place. This entails extracting relevant features such as bag-of-words representations, TF-IDF scores, or embeddings like Word2Vec from the preprocessed text. The next step is to select an appropriate model for sentiment analysis. This can range from traditional machine learning algorithms like Naive Bayes and Support Vector Machines to advanced deep learning models like Recurrent Neural Networks (RNNs) or Transformer-based architectures such as BERT. Training the selected model involves dividing the data into training and testing sets and then fine-tuning and optimizing it to improve performance. Evaluation metrics such as accuracy, precision, and recall are used to determine the effectiveness of the model. Once satisfactory performance is achieved, the model can be deployed for real-time sentiment analysis either through API integration or web application deployment. Ongoing monitoring and maintenance are essential to ensure that the model remains accurate and up-to-date with evolving language patterns and sentiments on Twitter. Moreover, ethical considerations must be taken into account throughout the development and deployment process to protect privacy and mitigate bias. By adopting this systematic approach towards harnessing Twitter for automatic sentiment identification, it becomes an invaluable resource for applications in market research, brand monitoring, social media analytics amongst others.

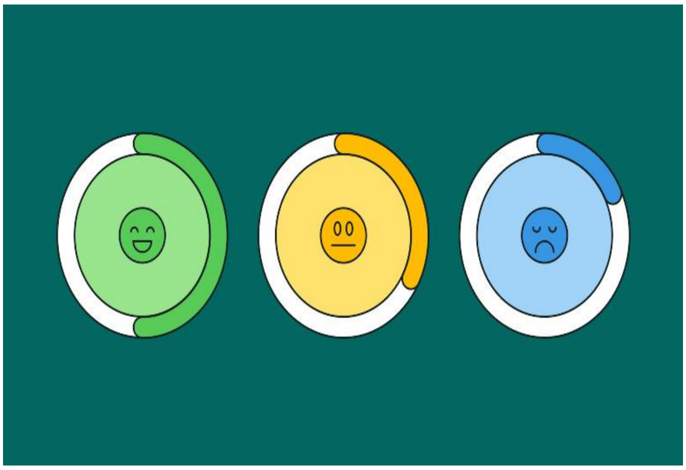


Figure-1: Sentiment Identification of Social Media Post.

1.1. Purpose of Sentiment Identification

Sentiment identification is a crucial tool in analyzing Twitter data as it serves multiple purposes. Firstly, it provides invaluable insights into consumer perceptions, which is essential for companies to understand customer sentiment towards their products or services. By discerning positive, negative, or neutral sentiments from tweets, businesses can tailor their strategies to meet customer needs effectively. This can lead to improved customer satisfaction and loyalty.

Secondly, sentiment analysis enables proactive brand monitoring, allowing companies to track their brand reputation in real-time and address any emerging issues promptly. This is particularly important in today's digital age where news spreads rapidly online. By detecting negative sentiment early on, companies can take swift action to safeguard their brand reputation and foster positive stakeholder engagement.Moreover, sentiment identification aids in extracting actionable feedback from customer interactions on Twitter. This feedback can be used to continuously improve products or services and enhance customer satisfaction. It also facilitates competitive analysis by evaluating how competitors are perceived on social media platforms like Twitter. By identifying areas of differentiation, businesses can make strategic decisions that give them a competitive edge.

Sentiment identification plays a crucial role in crisis management by detecting and addressing negative sentiment early on. This helps safeguard brand reputation during times of crisis when emotions run high online. By addressing issues promptly and transparently, companies can turn potential crises into opportunities for growth and improvement. Sentiment identification on Twitter serves as a cornerstone for data-driven decision-making, brand management, and maintaining a competitive edge in today's dynamic business landscape. Its ability to provide insights into consumer perceptions, enable proactive brand monitoring, extract actionable feedback from customer interactions on Twitter, facilitate competitive analysis, trend spotting and crisis management makes it an essential tool for any business looking to succeed in the digital era.

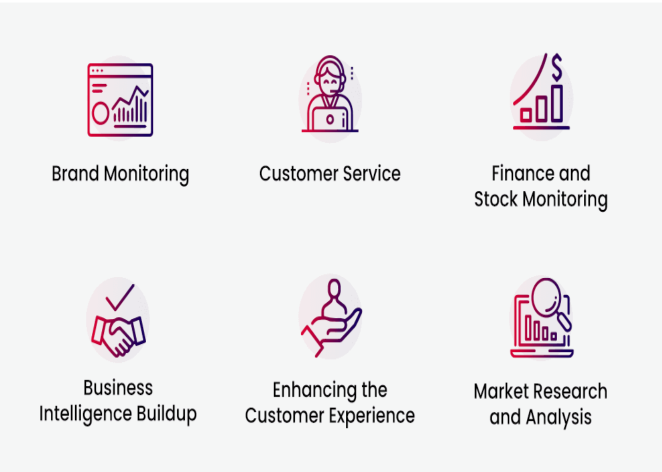


Figure-2: Purpose of Sentiment Identification.

1.2.Obstacles Encountered in Sentiment Analysis.

Sentiment analysis is a valuable tool for comprehending public opinion and customer sentiment. However, this analytical technique faces significant difficulties in accurately interpreting and analyzing text data. One of the main issues is the inherent ambiguity and complexity of language, where identical words or phrases might convey different sentiments depending on the context. Additionally, expressions like sarcasm, irony, and humor pose substantial challenges for sentiment analysis models, frequently resulting in incorrect interpretations of sentiment. Negation words and modifiers further complicate the process by altering the polarity of sentiments within sentences.

Furthermore, subjectivity in sentiment perception, data sparsity, and imbalance in labeled datasets present significant obstacles in training accurate sentiment analysis models. Domain adaptation is also an issue, particularly in specialized domains where sentiment expressions may differ widely across various contexts. The multilingual nature of text data and temporal dynamics of sentiment expression on social media platforms like Twitter add further complexity to sentiment analysis tasks.

To overcome these difficulties requires constant research and development of advanced natural language processing techniques that can enhance the accuracy, robustness, and adaptability of sentiment analysis models across diverse languages, domains, and temporal contexts. While this may be a challenging endeavor, it is essential for achieving reliable insights into public opinion and customer feedback that can inform critical decision-making processes.

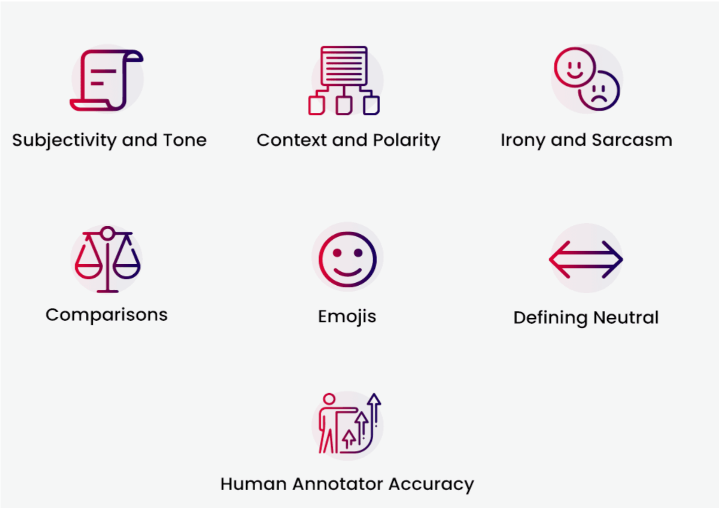


Figure-3: Obstacles Encountered in Sentiment Analysis.

2. API OF SOCIAL MEDIA

Social media APIs are an essential tool for developers, as they provide a versatile set of tools and protocols to interact programmatically with major platforms such as Facebook, Twitter, Instagram, LinkedIn, and YouTube. These APIs allow developers to seamlessly integrate social media functionality into their applications or create custom tools for managing social media presence. With the help of these APIs, developers can perform various tasks like posting content, retrieving user data, analyzing engagement metrics, managing advertising campaigns, and moderating content.

To interact with these APIs effectively, developers leverage HTTP requests and software development kits (SDKs) provided by the platforms. However, it is crucial to adhere to each platform's terms of service and usage policies while integrating these APIs. Developers should also be mindful of rate limits and usage restrictions imposed by social media APIs to ensure fair and responsible integration.

Social media APIs offer a wide range of capabilities that enable developers to create innovative applications that enhance the user experience on popular platforms. These tools also provide a way for businesses to manage their social media presence effectively. By following best practices and adhering to guidelines set forth by each platform provider, developers can integrate social media functionality into their applications seamlessly and responsibly.

2.1.Condensed: Get Twitter data with Twython API.

To effectively determine the sentiment (positive, negative, or neutral) of tweets regarding a specific product or movie, it is crucial to gather only those tweets that are directly related to the subject matter. The main goal of the thesis is to analyze the sentiments expressed in tweets that are relevant to the product or movie being studied. This requires scrutinizing the emotions and opinions conveyed in tweets that are pertinent to the topic at hand. However, this task is not without its challenges, as there is no surefire way to capture all tweets pertaining to a particular subject. Although there may be obstacles in the way, it remains crucial to gather a wide range of relevant tweets in order to carry out an accurate sentiment analysis and draw significant conclusions from the gathered data. As such, it is paramount that researchers take a meticulous approach when selecting tweets for their analysis to ensure that their results are both reliable and insightful. By carefully scrutinizing each tweet, researchers can weed out irrelevant or misleading information and focus solely on those tweets that hold true value and meaning. This level of attention to detail will ultimately lead to more accurate and meaningful findings, providing valuable insights into the sentiments of Twitter users on a particular topic or issue.

```python
from twython import Twython

# Replace these values with your own Twitter API credentials
APP_KEY = 'your_app_key'
APP_SECRET = 'your_app_secret'
ACCESS_TOKEN = 'your_access_token'
ACCESS_TOKEN_SECRET = 'your_access_token_secret'

# Initialize Twython with your API credentials
twitter = Twython(APP_KEY, APP_SECRET, ACCESS_TOKEN, ACCESS_TOKEN_SECRET)

# Post a tweet
twitter.update_status(status='Hello, Twitter API! This tweet was posted using Twython

# Fetch user data
user_data = twitter.show_user(screen_name='twitter')
print("User Details:")
print("Name:", user_data['name'])
print("Screen Name:", user_data['screen_name'])
print("Followers Count:", user_data['followers_count'])
print("Following Count:", user_data['friends_count'])
```

Figure-4: Condensed: Get Twitter data with Twython API.

3.PROCESS OF REMOVING HASH SYMBOL FROM SENTIMENT POST FROM SOCIAL MEDIA

If you're looking to eliminate the hashtag symbol (#) from a sentiment post on any social media platform, there are some simple steps you can take. First and foremost, locate the post that contains the hashtag you want to remove. Once you've found it, click on the three dots or options icon usually located in the top right-hand corner of the post. From there, select the option to edit your post. This will allow you to modify your post's content and delete any hashtags that are present. After making your desired changes, save your updated post by clicking "save" or "update." It's important to note that removing a hashtag from a sentiment post may impact its visibility and reach among other users on the platform. However, if you feel that it is necessary to remove the hashtag for personal or professional reasons, these steps should help you do so easily and effectively.

- **Identify the sentiment post:** Locate the post containing the hash symbol (#) that you want to remove.
- **Copy the post:** Highlight the text of the sentiment post, including the hash symbol, and copy it.
- **Open a text editor or word processor:** Open a text editor or word processor application on your computer or device.
- **Paste the copied text:** Paste the sentiment post into the text editor or word processor.
- **Find and replace:** Use the find and replace function (usually accessible through a menu or keyboard shortcut) to replace all instances of the hash symbol (#) with an empty space or any other character you prefer.
- **Review the modified post:** Check the modified sentiment post to ensure that the hash symbols have

been removed correctly and that the sentiment post still makes sense.

- **Save the modified post:** Once you're satisfied with the changes, save the modified sentiment post.
- **Post the modified sentiment:** If you intend to repost the sentiment on social media, copy the modified text from the text editor or word processor and paste it into the social media platform of your choice.

4.TYPE OF SENTIMENT ANALYSIS OF SOCIAL MEDIA POST

The process of analyzing sentiments in social media posts can be broadly categorized into several types, depending on the scope, method, and purpose of the analysis. There are four widely recognized types that are commonly employed in this regard. The first type is known as document-level sentiment analysis, which involves analyzing the sentiments expressed in individual social media posts or documents. The second type is aspect-based sentiment analysis, which focuses on identifying the sentiment associated with specific aspects or features of a product or service mentioned in social media posts. The third type is domain-based sentiment analysis, which involves analyzing sentiments across different domains or topics, such as politics, sports, or entertainment. Finally, there is cross-lingual sentiment analysis that aims to identify and analyze sentiments expressed in different languages. These are just some examples of the various types of sentiment analysis that can be applied to social media data to gain valuable insights into customer opinions and preferences.

Basic Sentiment Analysis: This type of sentiment analysis categorizes text into positive, negative, or neutral sentiments. It usually involves using lexicons or machine learning algorithms to classify the sentiment polarity of individual posts.

Aspect-Based Sentiment Analysis: Aspect-based sentiment analysis goes beyond just determining the overall sentiment of a text and identifies the sentiment towards specific aspects or entities mentioned within the text. For example, in a product review, it can analyze sentiment towards various features or attributes of the product separately.

Emotion Detection: Emotion detection aims to identify specific emotions expressed in social media posts, such as joy, anger, sadness, or fear. This type of sentiment analysis often involves more sophisticated natural language processing techniques, including deep learning models trained specifically for emotion recognition.

Opinion Mining: Opinion mining, also known as subjectivity analysis, involves identifying not just sentiment but also the opinions or attitudes expressed in social media posts. It aims to understand the stance or viewpoint of the author towards a particular topic or entity.

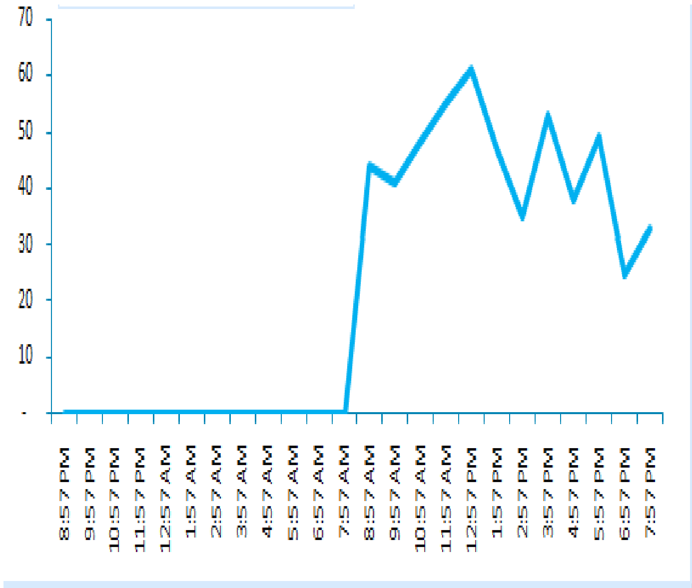
5.ANALYSIS OF SENTIMENT IDENTIFICATION OF POST

Sentiment identification of posts involves a multi-step process to analyze textual data and discern the emotional tone conveyed by the author. Initially, the text undergoes preprocessing steps such as tokenization, lowercasing, and removal of stopwords, punctuation, and special characters to clean the data for analysis. Following this, there are two primary approaches employed: lexicon-based methods and machine learning-based methods. Lexicon-based methods rely on sentiment lexicons or dictionaries containing lists of words associated with different sentiment scores, while machine learning-based methods train models on labeled data to predict sentiment. Features are then extracted from the text data, utilizing techniques such as bag-of-words, TF-IDF, or word embeddings. These features are used to train and evaluate machine learning models on labeled datasets, assessing their performance using metrics like accuracy and F1-score. In the world of natural language processing, sentiment analysis is an essential task that involves identifying and categorizing the emotions expressed in textual data. To achieve this goal, a comprehensive process is required, which includes several steps such as text preprocessing, feature extraction, model training, and evaluation techniques.

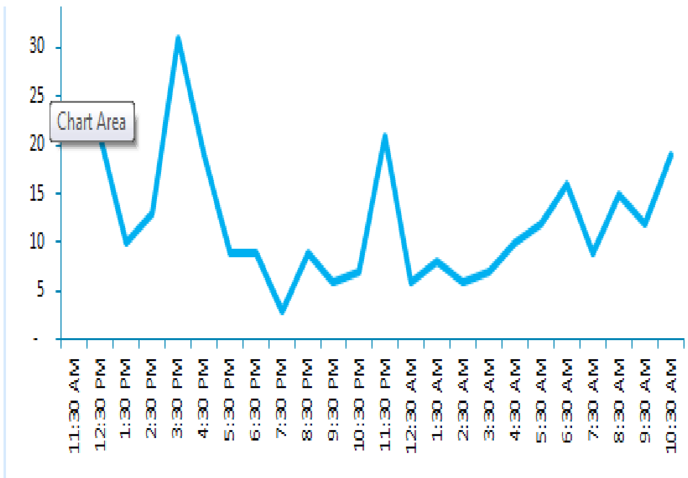
One of the most popular approaches to sentiment analysis is using deep learning models like BERT or GPT. However, these models need to be fine-tuned to adapt them to the specific task of sentiment analysis. Once trained, these models can accurately predict the sentiment of new posts or text data.

After predicting the sentiment scores for each post or text data point, post-processing steps may be applied to classify them into positive, negative or neutral categories. These scores can then be aggregated for further analysis or visualization purposes.

Sentiment identification involves a rigorous process that requires expertise in different areas of natural language processing and machine learning. By combining various techniques and methods, it is possible to accurately determine the emotions expressed in textual data and gain valuable insights into people's opinions and attitudes towards different topics.



Graph-1: Summary: Count of tweets with hashtag "sad".



Graph-2: Tweet volume for "rage" at a glance.

6. DATASET OF EMOTION

The use of hashtags allows us to express our thoughts and feelings in a succinct and easily searchable manner. By simply adding a hashtag to our posts or messages, we can provide a wealth of information about what is on our minds, from the topics that interest us to the issues that concern us. Recognizing the immense value of this type of user-generated data, we have taken steps to enhance our machine learning algorithm to include hashtags in its analysis. This enables us to gain even deeper insights into people's thoughts and emotions, which then helps us better understand their needs and preferences.

Thanks to this upgraded algorithm, we are now able to process much larger amounts of data with greater accuracy and efficiency. As a result, we can improve our products and services by customizing them according to the specific needs of our customers. Ultimately, by harnessing the power of hashtags and other types of user-generated content, we can gain valuable insights into human behavior that enable us to create more personalized experiences for everyone. This is an exciting development that has far-reaching implications for businesses across all industries because it allows them to connect with their customers in ways they never could before. The future is bright for those who embrace these new technologies!

7. CONCLUSION

This academic research delves into the field of sentiment analysis and centers on the application of lexical resources and machine learning algorithms to classify the emotional tone of tweets and text messages, both of which are examples of unstructured data sources. With the vast amount of subjective information available online, Sentiment Analysis has become a valuable tool in various sectors such as online advertising and market research. In knowledge management, opinion data is a crucial factor that often determines significant decisions. The study aims to gain an understanding of the challenges associated with Sentiment Analysis and explore numerous approaches developed to tackle them. Extracting underlying sentiments from social media data can be a daunting task due to its volume and variety. To determine the most effective elements for Sentiment Analysis, we analyzed tweets from the public stream using both lexicon-based methods and machine-learning techniques. The results will provide insights into improving Sentiment Analysis performance in extracting subjective content from unstructured data sources.

REFERNCE

1. Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2):1–135, 2008.
2. Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd annual meeting on Association for Computational Linguistics, page 271. Association for Computational Linguistics, 2004.
3. Bing Liu. Opinion mining and sentiment analysis. In Web Data Mining, pages 459–526. Springer, 2011.
4. Bing Liu. Sentiment analysis and subjectivity. Handbook of natural language processing, 2:627–666, 2010.
5. Andr´es Montoyo, Patricio Mart´ıNez-Barco, and Alexandra Balahur. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. Decision Support Systems, 53(4):675–679, 2012.

6. Erik Cambria, Bjorn Schuller, Yunqing Xia, and Catherine Havasi. New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2):15–21, 2013.

7. Khairullah Khan, Baharum Baharudin, Aurnagzeb Khan, and Ashraf Ullah. Mining opinion components from unstructured reviews: A review. Journal of King Saud University-Computer and Information Sciences, 26(3):258–275, 2014.

8. Ronen Feldman, Moshe Fresko, Jacob Goldenberg, Oded Netzer, and Lyle Ungar. Extracting product comparisons from discussion boards. In Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on, pages 469–474. IEEE, 2007.

9. Mohammad Sadegh, Roliana Ibrahim, and Zulaiha Ali Othman. Opinion mining and sentiment analysis: A survey. International Journal of Computers & Technology, 2(3):171–178, 2012.

10. Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, pages 79–86. Association for Computational Linguistics, 2002.

11. Peter D Turney. 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, pages 417–424. Association for Computational Linguistics, 2002.

12. Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In Proceedings of LREC, volume 6, pages 417–422. Citeseer, 2006.

13. Alaa Hamouda and Mohamed Rohaim. Reviews classification using sentiwordnet lexicon. In World Congress on Computer Science and Information Technology, 2011

14. Doug Cutting, Julian Kupiec, Jan Pedersen, and Penelope Sibun. A practical part-of-speech tagger. In Proceedings of the third conference on Applied natural language processing, pages 133–140. Association for Computational Linguistics, 1992.

15. Shitanshu Verma and Pushpak Bhattacharyya. Incorporating semantic knowledge for sentiment analysis. Proceedings of ICON, 2009.

16. Kamal Nigam, John La_erty, and Andrew McCallum. Using maximum entropy for text classification In IJCAI-99 workshop on machine learning for information filtering, volume 1, pages 61–67, 1999.

17. Huifeng Tang, Songbo Tan, and Xueqi Cheng. A survey on sentiment detection of reviews. Expert Systems with Applications, 36(7):10760–10773, 2009.

18. Daniel M Bikel and Je_rey Sorensen. If we want your opinion. In Semantic Computing, 2007. ICSC 2007. International Conference on, pages 493–500. IEEE, 2007

19. Hiroshi Kanayama and Tetsuya Nasukawa. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pages 355–363. Association for Computational Linguistics, 2006.

20. Matthias Hagen, Martin Potthast, Michel B¨uchner, and Benno Stein. Twitter sentiment detection via ensemble classification using averaged confidence scores. In Advances in Information Retrieval, pages 741–754. Springer, 2015.

21. Saif M Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. 2013.