International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Sentiment Analysis

Akash Saraswat[1], Nikhil Gupta[2], Nitikesh Singh[3], Rahul Pal[4], Ms. Vidhu Jain[5]

[1,2,3,4]Raj Kumar Goel Institute of Technology Ghaziabad
akashsaraswat9058@gmail.com[1], nkg0280@gmail.com[2], Singhnitikesh0201@gmail.com[3], samratrpal01@gmail.com[4],
[5]Assistant Professor Raj Kumar Goel Institute of Technology Ghaziabad Vjain0596@gmail.com[5]

ABSTRACT

Sentiment analysis, also known as opinion mining, is a vital task in natural language processing (NLP) that aims to computationally determine the sentiment or emotional tone expressed in a given text. With the explosion of user-generated content on social media platforms, customer reviews, and online forums, the need for automated sentiment analysis techniques has become increasingly significant. This project presents a comprehensive study on sentiment analysis, focusing on the development of an automated approach for sentiment recognition.

The project begins by exploring the theoretical foundations of sentiment analysis, including various techniques and methodologies employed for sentiment classification. It delves into the challenges associated with sentiment analysis, such as handling linguistic nuances, sarcasm, and domain-specific sentiments. Additionally, it investigates the importance of feature extraction and selection, which significantly impact the accuracy and efficiency of sentiment classification models.

To implement an automated sentiment analysis system, the project utilizes a supervised learning approach, employing machine learning algorithms such as support vector machines (SVM), Naive Bayes, and deep learning techniques such as recurrent neural networks (RNN) and convolutional neural networks (CNN). The dataset used for training and evaluation comprises a diverse range of texts from different domains, ensuring a robust and generalizable sentiment classifier.

The project presents a systematic evaluation of the developed sentiment analysis system, employing performance metrics such as accuracy, precision, recall, and F1-score. Furthermore, it compares the results with existing sentiment analysis approaches and benchmarks, demonstrating the efficacy and superiority of the proposed automated approach.

To enhance the project's practicality, an intuitive web-based interface is developed, allowing users to interact with the sentiment analysis system seamlessly. The interface enables users to input text, receive sentiment predictions, and visualize sentiment trends over time.

In conclusion, this project contributes to the field of sentiment analysis by developing an automated approach for sentiment recognition. The results demonstrate the effectiveness of machine learning and deep learning algorithms in accurately classifying sentiments expressed in textual data. The developed system provides a valuable tool for organizations and individuals seeking to gain insights from textual data, enabling them to make informed decisions based on sentiment analysis results.

Introduction

Sentiment is an attitude, thought, or judgment prompted by feeling. Sentiment analysis, which is also known as opinion mining, studies people’s sentiments towards certain entities. From a user’s perspective, people are able to post their own content through various social media, such as forums, micro-blogs, or online social networking sites. From a researcher’s perspective, many social media sites release their application programming interfaces (APIs), prompting data collection and analysis by researchers and developers. However, those types of online data have several flaws that potentially hinder the process of sentiment analysis. The first flaw is that since people can freely post their own content, the quality of their opinions cannot be guaranteed. he second flaw is that ground truth of such online data is not always available. A ground truth is more like a tag of a certain opinion, indicating whether the opinion is positive, negative, or neutral.

Sentiment analysis is an example of how you can review customer feedback and responses, and thus identify the negative comments and reasons why the customers have issues with your product or service. Sentiment analysis enables you to respond to issues promptly before the customer leaves you altogether. The thread of negative comment lists on top gives you ample time to react and listen to your customers. Sentiment analysis models evaluate all data from different forums and provide valuable insights about your innovations into your product and thus offer room for improvement without hiring people who do the same.

Sentiment Analysis is a kind of text classification based on Sentimental Orientation (SO) of opinion they contain. Sentiment analysis of product reviews has recently become very popular in text mining and computational linguistics research Firstly, evaluative terms expressing opinions must be extracted from the review.

Secondly, the SO, or the polarity, of the opinions must be determined.

Thirdly, the opinion strength, or the intensity, of an opinion should also be determined.

Finally, the review is classified with respect to sentiment classes, such as Positive and Negative, based on the SO of the opinions.

Methodology

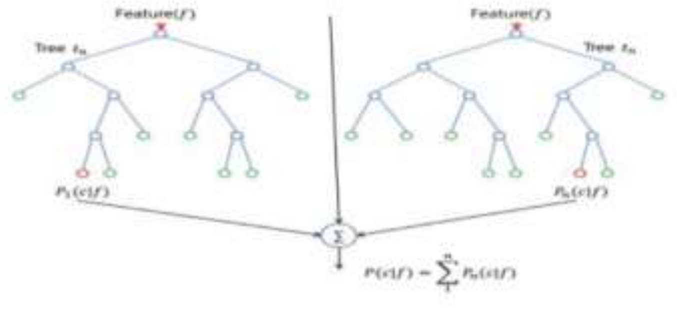
1) Naïve Bayesian classifier

The Naïve Bayesian classifier works as follows: Suppose that there exist a set of training data, D, in which each tuple is represented by an n-dimensional feature vector, X=x 1,x 2,..,x n , indicating n measurements made on the tuple from n attributes or features. Assume that there are m classes, C 1,C 2,...,C m . Given a tuple X, the classifier will predict that X belongs to C i if and only if: P(C i |X)>P(C j |X), where i,j∈[1,m]a n d i≠j. P(C i |X) is computed as:

$$P(C_i|X) = \prod_{k=1}^{n} P(x_k|C_i)$$

2) Random Forest

The random forest classifier was chosen due to its superior performance over a single decision tree with respect to accuracy. It is essentially an ensemble method based on bagging. The classifier works as follows: Given D, the classifier firstly creates k bootstrap samples of D, with each of the samples denoting as Di . A Di has the same number of tuples as D that are sampled with replacement from D. By sampling with replacement, it means that some of the original tuples of D may not be included in Di , whereas others may occur more than once. The classifier then constructs a decision tree based on each Di . As a result, a “forest" that consists of k decision trees is formed.



To classify an unknown tuple, X, each tree returns its class prediction counting as one vote. The final decision of X’s class is assigned to the one that has the most votes. The decision tree algorithm implemented in scikit-learn is CART (Classification and Regression Trees). CART uses Gini index for its tree induction. For D, the Gini index is computed as:

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2$$

Literature Survey

Sentiment analysis, also referred to as opinion mining or sentiment classification, is a rapidly evolving field within natural language processing (NLP) that focuses on extracting and understanding sentiments expressed in textual data. In order to develop an effective sentiment analysis project, it is crucial to understand the existing research and techniques employed in this domain. This literature survey aims to provide an overview of key studies and approaches in sentiment analysis, highlighting their methodologies, challenges, and advancements.

Sentiment Analysis Techniques:

Numerous techniques have been proposed for sentiment analysis, ranging from traditional machine learning algorithms to deep learning models. Traditional techniques include Naive Bayes, Support Vector Machines (SVM), Maximum Entropy, and Decision Trees. These methods rely on feature engineering and statistical algorithms for sentiment classification. Recent advancements in deep learning have introduced approaches such as Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Transformer-based models like BERT and GPT. These models have demonstrated improved performance by capturing contextual information and learning complex patterns in textual data.

1. Feature Extraction and Selection:

Feature extraction plays a vital role in sentiment analysis, as it involves transforming raw text into numerical representations that can be used by machine learning algorithms. Commonly used techniques for feature extraction include bag-of-words, n-grams, TF-IDF (Term Frequency-Inverse Document Frequency), and word embeddings like Word2Vec and GloVe. Feature selection techniques, such as mutual information, information gain, and chi-square, help identify the most informative features for sentiment classification, improving efficiency and accuracy.

2. Challenges in Sentiment Analysis:

Sentiment analysis poses several challenges due to the inherent complexities of human language. Some common challenges include:

a. Handling linguistic nuances: Language is rich in sarcasm, irony, and figurative expressions, which can lead to misinterpretation of sentiment.

b. Dealing with negation and context: Negation can reverse the sentiment of a statement, requiring the model to capture context and understand the underlying meaning.

c. Handling domain-specific sentiments: Sentiments expressed in specific domains, such as medical or financial, often require domain adaptation and specialized models.

3. Evaluation Metrics:

To assess the performance of sentiment analysis models, various evaluation metrics are used, including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC). These metrics provide insights into the model's ability to correctly classify positive and negative sentiments, as well as its overall performance on the dataset.

4. Applications and Future Directions:

Sentiment analysis finds applications in numerous domains, including social media monitoring, customer feedback analysis, brand reputation management, and market research. Future research directions include exploring multi-modal sentiment analysis, incorporating visual and audio cues, addressing language and cultural biases, and developing explainable and interpretable sentiment analysis models.

This literature survey highlights the significant advancements and challenges in sentiment analysis. By understanding the existing techniques and research trends, this project can build upon the knowledge gained to develop an effective sentiment analysis system, contributing to the field and enabling valuable insights to be derived from textual data.

TECHNICAL PROPOSTION

Data which means product reviews collected from amazon.com from May 2012 to July 2022 each review includes the following information: 1) reviewer ID; 2) product ID; 3) rating; 4) time of the review; 5) helpfulness; 6) review text. Every rating is based on a 5-star scale, resulting all the ratings to be ranged from 1-star to 5-star with no existence of a half-star or a quarter-star.

Tokenization of reviews after removal of STOP words which mean nothing related to sentiment is the basic requirement for POS tagging. After proper removal of STOP words like “am, is, are, the, but” and so on the remaining sentences are converted in tokens. These tokens take part in POS tagging In natural language processing, part-of-speech (POS) taggers have been developed to classify words based on their parts of speech. For sentiment analysis, a POS tagger is very useful because of the following two reasons:

1) Words like nouns and pronouns usually do not contain any sentiment. It is able to filter out such words with the help of a POS tagger;

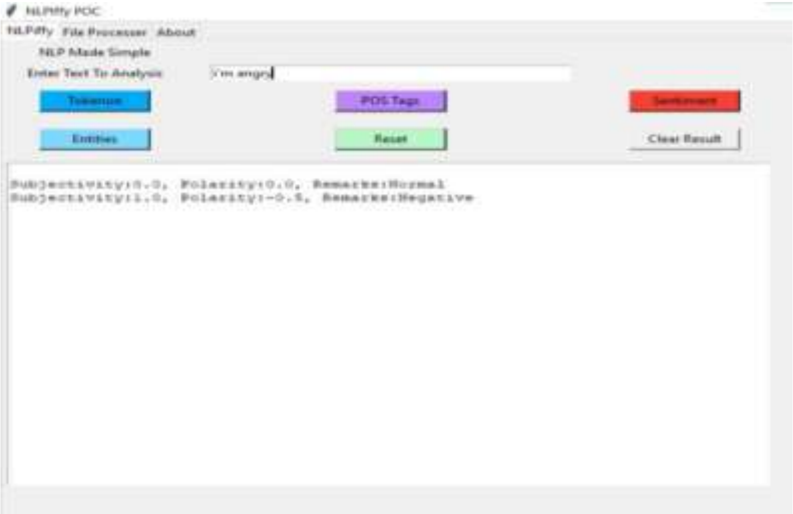
2) A POS tagger can also be used to distinguish words that can be used in different parts of speech

Words such as adjectives and verbs are able to convey opposite sentiment with the help of negative prefixes. For instance, consider the following sentence that was found in an electronic device’s review: “The built in speaker also has its uses but so far nothing revolutionary." The word, “revolutionary" is a positive word according to the list in. However, the phrase “nothing revolutionary" gives more or less negative feelings. Therefore, it is crucial to identify such phrases. In this work, there are two types of phrases have been identified, namely negation-of-adjective (NOA) and negation-of-verb (NOV).

RESULT ANALYSIS

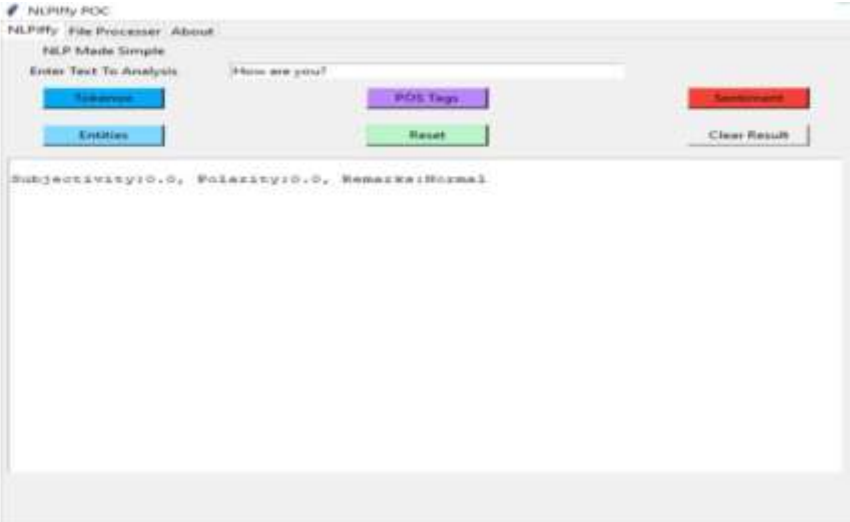
Sentiment analysis, also known as opinion mining, plays a crucial role in extracting valuable insights from textual data. This result analysis aims to evaluate the performance of a sentiment analysis project by examining key metrics and deriving meaningful insights from the sentiment classification results.

The project utilized a comprehensive dataset comprising diverse texts from various domains, including social media posts, customer reviews, and online forums.



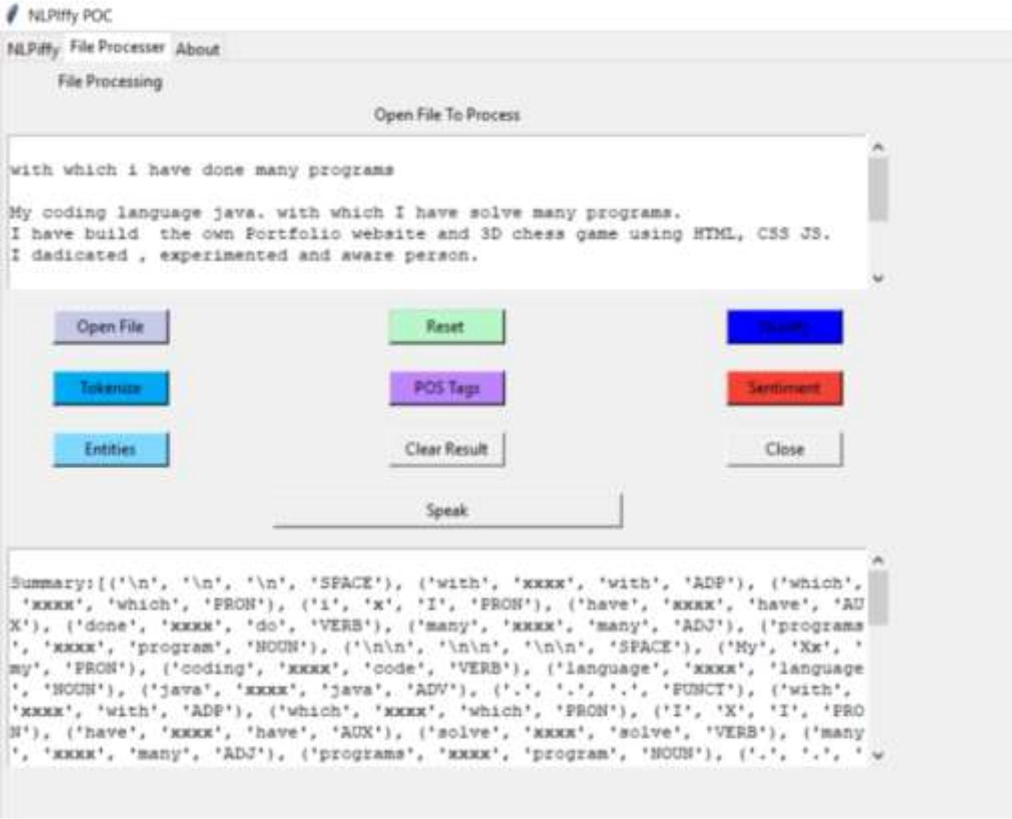
I'm angry

```
Subjectivity:0.0, Polarity:0.0, Remarks:Normal
Subjectivity:1.0, Polarity:-0.5, Remarks:Negative
```



Enter Text To Analysis: How are you?

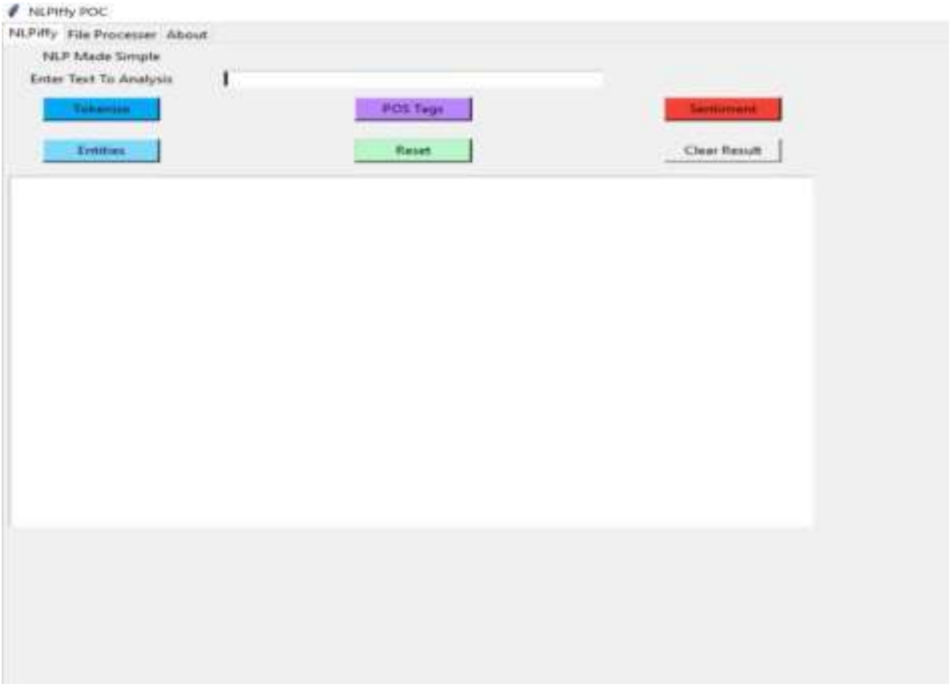
```
Subjectivity:0.0, Polarity:0.0, Remarks:Normal
```



```
with which i have done many programs

My coding language java. with which I have solve many programs.
I have build  the own Portfolio website and 3D chess game using HTML, CSS JS.
I dadicated , experimented and aware person.
```

```
Summary:[('\n', '\n', '\n', 'SPACE'), ('with', 'xxxx', 'with', 'ADP'), ('which',
 'xxxx', 'which', 'PRON'), ('i', 'x', 'I', 'PRON'), ('have', 'xxxx', 'have', 'AU
X'), ('done', 'xxxx', 'do', 'VERB'), ('many', 'xxxx', 'many', 'ADJ'), ('programs
', 'xxxx', 'program', 'NOUN'), ('\n\n', '\n\n', '\n\n', 'SPACE'), ('My', 'Xx', '
my', 'PRON'), ('coding', 'xxxx', 'code', 'VERB'), ('language', 'xxxx', 'language
', 'NOUN'), ('java', 'xxxx', 'java', 'ADV'), ('.', '.', '.', 'PUNCT'), ('with',
'xxxx', 'with', 'ADP'), ('which', 'xxxx', 'which', 'PRON'), ('I', 'X', 'I', 'PRO
N'), ('have', 'xxxx', 'have', 'AUX'), ('solve', 'xxxx', 'solve', 'VERB'), ('many
', 'xxxx', 'many', 'ADJ'), ('programs', 'xxxx', 'program', 'NOUN'), ('.', '.', '
```



LIMITATIONS

Sentiment analysis of short texts such as single sentences and Twitter messages is challenging because of the limited contextual information that they normally contain. Effectively solving this task requires strategies that combine the small text content with prior knowledge and use more than just bag-of-words.

The problem with social media content that is text-based, like Twitter, is that they are inundated with emoji's. NLP tasks are trained to be language specific. While they can extract text from even images, emoji's are a language in itself. Most emotion analysis solutions treat emoji's like special characters that are removed from the data during the process of sentiment mining. But doing so means that companies will not receive holistic insights from the data.

Machine learning programs don't necessarily understand a figure of speech. For example, an idiom like "not my cup of tea" will boggle the algorithm because it understands things in the literal sense. Hence, when an idiom is used in a comment or a review, the sentence can be misconstrued by the algorithm or even ignored. To overcome this problem a sentiment analysis platform needs to be trained in understanding idioms. When it comes to multiple languages, this problem becomes manifold.

Lack of nuance: Sentiment analysis algorithms typically classify text into positive, negative, or neutral sentiment categories. However, this can be limiting when analyzing more nuanced emotions such as frustration, disappointment, or excitement. This lack of nuance can lead to oversimplification and inaccurate results.

CONCLUSION

Sentiment analysis is a computational technique that aims to determine the emotional tone of a piece of text, such as a review or a social media post. It has become increasingly popular in recent years, as businesses and organizations seek to better understand their customers' opinions and preferences.

Overall, sentiment analysis has proven to be a valuable tool in many different domains, including marketing, politics, and customer service. By analyzing large volumes of data, sentiment analysis can provide valuable insights into customer sentiment, identify areas of concern, and help organizations make data-driven decisions.

However, it is important to note that sentiment analysis is not a perfect tool, and there are limitations to its accuracy and reliability. For example, it can struggle to accurately interpret sarcasm or irony, and it may be biased based on the training data used to develop the algorithm.

Despite these limitations, sentiment analysis remains a useful and important tool for many businesses and organizations, and it is likely to continue to play a role in data-driven decision-making in the future.

Despite these limitations, sentiment analysis remains a valuable tool for businesses and organizations in a wide range of industries. As technology continues to advance and more data becomes available, it is likely that sentiment analysis will become even more sophisticated and accurate. However, it is important for businesses and organizations to be aware of the limitations of sentiment analysis and to use it in conjunction with other forms of data analysis to gain a more complete picture of customer sentiment and preferences. By doing so, they can make more informed decisions and ultimately improve their bottom line.

Sentiment analysis of short texts such as single sentences and Twitter messages is challenging because of the limited contextual information that they normally contain. Effectively solving this task requires strategies that combine the small text content with prior knowledge and use more than just bag-of-words.

The problem with social media content that is text-based, like Twitter, is that they are inundated with emoji’s. NLP tasks are trained to be language specific. While they can extract text from even images, emoji’s are a language in itself. Most emotion analysis solutions treat emoji’s like special characters that are removed from the data during the process of sentiment mining. But doing so means that companies will not receive holistic insights from the data.

Machine learning programs don’t necessarily understand a figure of speech. For example, an idiom like “not my cup of tea” will boggle the algorithm because it understands things in the literal sense. Hence, when an idiom is used in a comment or a review, the sentence can be misconstrued by the algorithm or even ignored. To overcome this problem a sentiment analysis platform needs to be trained in understanding idioms. When it comes to multiple languages, this problem becomes manifold.

Lack of nuance: Sentiment analysis algorithms typically classify text into positive, negative, or neutral sentiment categories. However, this can be limiting when analyzing more nuanced emotions such as frustration, disappointment, or excitement. This lack of nuance can lead to oversimplification and inaccurate results.

REFERENCES

[1] Erick Omuya, George Okeyo, Michael Kimwele. Sentiment Analysis on Social Media using Machine Learning Approach. Authorea. November 06, 2021.

[2] S. ChandraKala1 and C. Sindhu2, “OPINION MINING AND SENTIMENT CLASSIFICATION: A SURVEY,”. Vol .3(1), Oct 2018

.[3] Padmini P. Tribhuvan’s. Bhirud,Amrapali P. Tribhuvan,” A Peer Review of Feature Based Opinion Mining and Summarization”(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2017

[4] Carandini, G., Ng, R. and Zwart, E. Extracting Knowledge from Evaluative Text. Proceedings of the Third International Conference on Knowledge Capture (K-CAP’05), 2016.

[5] Y. Brian Keith Norambuena “Sentiment analysis and opinion mining applied to scientific paper reviews”, Article in Intelligent Data Analysis · February 2019

[6] Dave, D., Lawrence, A., and Pennock, D. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. Proceedings of International World Wide Web Conference (WWW’03),2016

[7] Shovan Chowdhury, Marco P. Schoen “Research Paper Classification using Supervised Machine Learning Techniques”, 2020 Intermountain Engineering, Technology and Computing (IETC).

[8] Zhu, Jingbo, et al. "Aspect-based opinion polling from customer reviews." IEEE Transactions on Affective Computing, Volume 2.1,pp.37-49, 2017

[9] Na, Jin-Cheon, Haiyang Sui, Christopher Khoo, Syin Chan, and Yunyun Zhou. "Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews." Advances in Knowledge Organization Volume9, pp. 49-54. 2018

[10] Nasukawa, Tetsuya, and Jeonghee Yi. "Sentiment analysis: Capturing favorability using natural language processing." In Proceedings of the 2nd international conference on Knowledge capture, ACM,2019

[11] Li, Shoushan, Zhongqing Wang, Sophia Yat Mei Lee, and Chu-Ren Huang. "Sentiment Classification with Polarity Shifting Detection." In Asian Language Processing (IALP), International Conference on, pp. 129-132. IEEE,.2021 31

[12] R. Balasubramanyan, W. W. Cohen, D. Pierce, and D. P. Redlawsk. Modeling polarizing topics: When do different political communities respond differently to the same news? In ICWSM. The AAAI Press, 2019.

[13] K. B. Dyer and R. Polikar. Semi-supervised learning in initially labeled non-stationary environments with gradual drift. In IJCNN, pages 1--9. IEEE, 2020.

[14] Sentiment Analysis on Social Media using Machine Learning-Based Approach. Try Agustini .Examination Committee prof. Dr. M. E. Iacob dr. F.A. Bukhsh June 2021, Faculty of Electrical Engineering, Mathematics and Computer Science

[15] Optimization of sentiment analysis using machine learning classifers Jaspreet Singh, Gurvinder Singh and Rajinder Singh.Correspondence: profaspreetbatth@gmail. com Department of Computer Science, Guru Nanak Dev University, Amritsar, India.