**Sentimental Analysis of World Current Affairs**

**Capstone Project Proposal**

**Submitted by:**

**(102003461) Pritika Sharma**

**(102003466) Sahil Kadiyan**

**(102183001) Abhinandan Singla**

**(102003066) Harishankar Kumar**

**BE Third Year- COE/ CSE**

**CPG No. \_\_\_\_\_**

Under the Mentorship of

Dr. Harkiran Kaur

Assistant Professor



**Computer Science and Engineering Department**

**Thapar Institute of Engineering and Technology, Patiala**

**February 2023**

**TABLE OF CONTENTS**

* Mentor Consent Form <Page No>
* Project Overview <Page No>
* Problem Statement …..
* Need Analysis (1 Page) …..
* Literature Survey (3 Pages) …..
* Objectives
* Methodology
* Project Outcomes & Individual Roles
* Work Plan
* Course Subjects
* References

**Mentor Consent Form**

I hereby agree to be the mentor of the following Capstone Project Team

|  |  |  |
| --- | --- | --- |
| **Project Title:** | | |
| **Roll No** | **Name** | **Signatures** |
| 102003461 | Pritika Sharma |  |
| 102003466 | Sahil Kadiyan |  |
| 102183001 | Abhinandan Singla |  |
| 102003066 | Harishankar Kumar |  |

NAME of Mentor: ……Dr. Harkiran Kaur… . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

SIGNATURE of Mentor: . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

NAME of Co-Mentor(if any): ……… .Dr. Ravneet Kaur . . . . . . . . . . . . . . . . . . . . . . . . .

SIGNATURE of Co-Mentor: . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

**Project Overview**

Nowadays, digital news serves a vital function in society by offering a source of information to the public. It serves as a means of communication that helps to educate and inform people about current events and issues. A large section of the public now follows internet sources rather than print or broadcast media for news and information purposes. Hence, social network analysis helps to analyze people’s opinions and sentiments on different topics.

Millions of people give their opinion on various topics on a daily basis on social media like Twitter, which has become one of the most popular social media platforms for the expression of thought, with 500 million tweets per day and around 259.4 million active users [1], according to the statistics.

Through capstone project, the aim is to analyze the current world affairs and their socio-economic impact on the world using tweets from Twitter. Through sentimental analysis, we aim to understand the trends that can reveal patterns and predict human behavior in co-relation with various domains which directly or indirectly impact the life of an ordinary person.

Using the tweets, we aim to apply Deep Learning and Natural Language Processing to perform trend analysis and sentimental analysis. By analyzing the emotional tone of the tweet, nations can re-strategize for the goodwill of the communities and peace by understanding the need and wishes of the world citizens for the greater good.

**Need Analysis**

With so much data online to evaluate, there is need for efficient ways to analyze and understand the opinions of the masses and their governing bodies. Sentiment analysis provides valuable insights into the public's attitudes and emotions toward various events and issues worldwide. Digital communication channels offer data rich with public opinion, perception, and potential risks related to world affairs.

By analyzing sentiment expressed in social media posts, news articles, and other sources, governments and national organizations can gain insights into how people feel about different events and issues and can also track public perceptions of government policies, international events, and other issues, which can aid in unanimous world peace with effective policies and strategies that align with public sentiment.

National organizations can gain insights into how people are likely to respond to future events or issues, allowing them to make informed decisions about policy, marketing, or other areas.

Identifying potential risks related to world affairs is a crucial benefit of sentiment analysis. By monitoring sentiment, organizations can identify potential issues and respond proactively to prevent them from becoming significant problems. Sentiment analysis can detect early warning signs of political unrest or identify public sentiment about a particular international issue.

**Literature Survey**

1) K.Subba Reddy, M.Surya Bhupal rao, Makineedi Raja Babu, P.Naveen Sundar Kumar (2023) in [2] implemented an intelligent and advanced data mining technique. The authors normalized the dataset and applied Hierarchical Tweet Expression Clustering (HTEC) algorithm, used to cluster the attributes according to distance. The authors applied Bag-of-Words (BOW) and Term Frequency- Inverse Document Frequency (IDM) for classifier training and testing operations. For predicting the sentiment of the emotional tone the authors used Spatial Dense Bi-directional Long Short Memory. The study concludes that proposed system can be an effective tool in terms of reduced computational burden and time consumption with optimal performance and efficiency in handling large dimensional data.

2) Ahmed, J., & Ahmed (2023)  in [3] analyze different machine learning techniques for fake news detection . Authors  used topic classification of News Articles using Scikit-Learn Python library's LabelEncoder. The Machine learning  Models used for sentiment analysis were Decision Tree, AdaBoost, K-Nearest Neighbour, and Random Forest Classifiers. The study concluded that different machine learning techniques can be used for fake news detection.

3) L. Almuqren and A. I. Cristea (2023) in [4] measured customer satisfaction toward Telecom company in Saudi Arabia,  monitoring in real-time their customers’ satisfaction on Twitter and from questionnaire analysis, and providing telecom companies’ recommendations based on monitoring real-time customers’ satisfaction through Twitter.

4) E. Sharma, A. Gaur and S. Singhal (2022) in [5] inspected adequacy of Support Vector Machine Classifier in Sentiment Analysis of India vs Pakistan T20 World Cup Match. They concluded that outcomes plainly show the dependence of SVM execution upon input dataset.

5) Swathi, T., Kasiviswanath, N. & Rao, A.A(2022) in [6]  provides a systematic review of the literature on stock price prediction using twitter sentiment analysis . Authors identifies an optimal deep learning technique named LSTM . The study concluded that LSTM can be an effective tool for stock prediction using twitter sentiment analysis.

6) C. Patra, P. S. D and S. Chakraborty (2022) in [7] analysed information flow of hashtag networks by combining sentiment analysis and graph techniques to look into the trending hashtag networks propagated by political parties using Twitter.

7) Demotte, P., Wijegunarathna, K., Meedeniya, D. *et al* (2021)in [8] produced efficient algorithms based on deep learning for text-based analytics. This study proposes an approach based on capsule networks for social media content analysis, especially for Twitter. The study concludes that the research findings indicate noteworthy accuracy enhancement for text processing within social media content analysis.

8) Gandhi, U.D., Malarvizhi Kumar, P., Chandra Babu, G. *et al* (2021)*.* in [9] perform  Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM).  The result of Deep Learning algorithms aims to rate the review tweets and also able to identify movie review.

9) J. K. Chandra, E. Cambria and A. Nanetti (2020) in [10] extracted sentiments of the Belt and Road Initiative from Twitter using aspect-based sentiment analysis. The authors employed a hybrid approach of symbolic and sub-symbolic techniques using gated convolutional networks, aspect embeddings and the SenticNet framework to solve the subtasks of aspect category detection and aspect category polarity

10) L. G. Singh, A. Anil and S. R. Singh (2020) in [11] proposed a semisupervised sentiment hashtag embedding (SHE) model, which is capable of preserving both semantic as well as sentiment distribution of the hashtags, SHE leverages a multitask learning approach using an autoencoder and a convolutional neural network-based classifier.

11) A. -L. Alten, G. Gadre, S. Kulkarni and C. -S. Wu (2019) in [12]  provides a systematic review of the literature on twitter sentiment analysis for Happines Index based on geographical location . The authors analysed polarity scoring using sentiment intensity. Visualization were done using D3 which is a javascript library. The study concluded  that Happiness Index can be calculated using Sentiment Intensity.

12) Herlawati, R. T. Handayanto, D. Setiyadi and E. Retnoningsih (2019) in [13] analyze how corpus can be used for twitter sentiment analysis. Authors used Naive bayes method for classification and various libraries were used for retrieving the tweets . The study concluded that naive bayes classifier can be an effective tool for sentiment analysis of tweets using corpus.

13) V. Rekha, R. Raksha, P. Patil, N. Swaras and G. L. Rajat (2019) in [14] discovered different Indian government schemes that was prepared and analysed through machine learning. the cleaned data was applied on three machine learning classifiers, namely, naïve bayes, random forest and tailored random forest. Random forest provided better accuracy in sentiment prediction than naïve bayes. Tailored random forest involved adding few extra parameters which produced an observable increment in the accuracy of prediction.

14) A. Aslam (2019) , .. in [15] formulated a structured way to extract opinions from twitter by  the use of sentiment analysis, tokenization and Average TF-IDF. The study concludes that average TF-IDF can be an effective tool for extracting opinions from twitter.

15) G. Yang, H. He and Q. Chen(2019) in [16] interpret emoticons not as grammatical functions but as an emotional signal in an independent form for sentimental analysis. Authors analyzed different techniques for emotion semantic analysis . Different techniques includes ECNN model , SVM for emotional classification ,Bayesian model and enhanced neural network . The study concluded that enhanced neural network can be an effective tool for emoticons sentiment analysis

16) S. Dorle and N. Pise (2018) in [17] achieved tweets analysis using RNN models. The authors did so by analysing streams of data containing directed loops  representing  the propagation of activations to future inputs which   accept the features for processing and training was done using gradient descent approach. The study concludes that RNN can be an effective model for sentiment analysis.

17) S. Wankhede, R. Patil, S. Sonawane and P. A. Save (2018) in [18] analyzed  tweets with emoticons and spell-checking and  correction of tweets. Authors calcaultes the sentiment score using content classifier and emoji sentiment  ranking was used for ranking the emoticons of sentiments which uses mapping of emoji . N-gram model was used for spelling detection and for the correction Hidden markov model(HMM) was used . The study concluded that N-gram can be an effective tool for data preprocessing for efficient sentiment analysis.

18) S. Geetha and V. K. Kaliappan (2018) in [19] predicted different texts formats from twitter, by considering text context associated with the emoticons and punctuations, the authors devised novel Future Prediction Architecture Based On Efficient Classification (FPAEC), designed with various classification algorithms such as, Fisher's Linear Discriminant Classifier (FLDC), Support Vector Machine (SVM), Naïve Bayes Classifier (NBC) and Artificial Neural Network (ANN) Algorithm along with the BIRCH

19) J. Chen, H. Li, Z. Wu and M. S. Hossain (2017)in [20] proposed a method to calculate the sentiment score for each tweet and each Twitter user, they adopted correlation coefficient between the sentiment scores of regular tweets and those of retweets to measure the influence.

**Objectives**

Through this capstone project, we aim to achieve the following:

* Classifying the tweets into sub topics using topic modeling algorithms.
* To Identify the overall sentiments revealing whether public opinion towards the current world affair is positive, negative, or neutral concerning various ethnicity and national divisions.
* To compare and analyze the emotional tone based on temporal effect. That is during specific events such as the pre-crisis, post-crisis, and period of crisis.
* To geographically analyze the regional differences in sentiments and approaches towards the current world affairs, helping to identify the critical areas of potential conflict or cooperation.
* To identify Social Butterflies such as influential individuals, national and international organizations, or media sources that shape public opinion towards current issues.

**Methodology**

The method through which we aim to analyze the data is by breaking the process into different stages, which include data collection, data preprocessing, data clustering, trend analysis, sentimental analysis, opinion mining, and interpretation and visualization. These are discussed below:

**Data Collection:** Collect data from various sources, such as social media platforms and other relevant textual sources, using NLP libraries and tools.

**Data Pre-processing:** Clean and pre-process the retrieved data, including removing irrelevant information, formatting, and text normalization.

**Data Clustering:** Classification of tweets under various domains of interest and popularity using various clustering techniques.

**Trend Analysis:** Analyze the data to identify trends, patterns, and topics most relevant to the research objective. Use topic modelling, keyword extraction, and clustering techniques to identify the most pertinent issues.

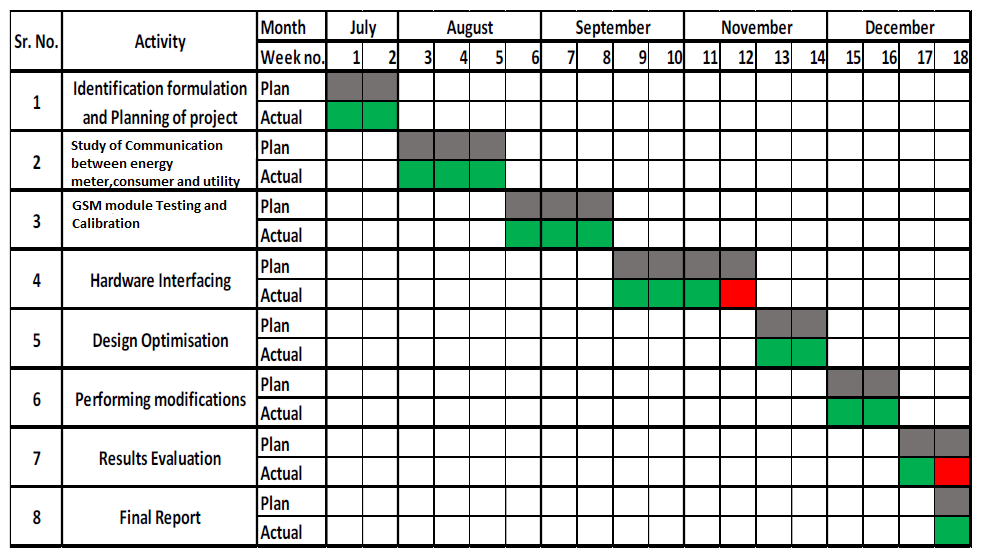
**Sentiment Analysis:** Apply sentiment analysis techniques to the data to identify the sentiment towards the topics of interest. Using machine learning algorithms to classify the sentiment of the text as positive, negative, or neutral.

**Opinion Mining:** Identify the opinions and attitudes of Twitter users towards the topics of interest. Which involves analyzing the text to identify the beliefs, attitudes, and opinions.

**Interpretation and Visualization:** Analyze the results to identify the most pressing issues and insights with visualization.

**Work Plan**

* You need to give a short work plan which you will set for achieving the set objectives, in this section.
* Sample:-



**Project Outcomes & Individual Roles**

* **Topic wise classification and popularity of tweets.**
* **Identification of overall sentiment:** The sentiment analysis would reveal whether public opinion towards the current world affair is positive, negative, or neutral concerning ethnicity.
* **Temporal comparison of sentiments:** It would depict the shift in the emotional tone of the people over time and during specific events such as the pre-crisis, post-crisis, and period of crisis.
* **Geographical analysis:** It would present the regional differences in sentiment and approaches towards the current world affairs, helping to identify the critical areas of potential conflict or cooperation.
* **Identification of influential actors/ Social Butterflies:** It would help identify the influential individuals, national and international organizations, or media sources shaping public opinion toward the topic in trend.

**Roles of each member:**

1. Pritika Sharma- Implementation of state of art DL models for data analysis and clustering for popularity analysis.
2. Sahil Kadiyan- Retrieving data using various web scrapping tools and Twitter API and performing sentimental analysis.
3. Harishankar Kumar- Analyzing the preprocessed data using ML models with reference to the temporal and demography.
4. Abhinandan Singla- Trend analysis of the retrieved data using NLP models using hashtags and identification of social butterflies .

**Course Subjects**

The subjects mentioned below in the table 1.0 would be used during the successful execution of the capstone project.

**Table 1.0**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Subject Code** | **Subject Name** |
| 1 | UCS664 | Natural Language Processing |
| 2 | UML501 | Machine Learning |
| 3 | UCS411 | Artificial Intelligence |
| 4 | UCS503 | Software Engineering |
| 5 | UTA025 | Innovation and Entrepreneurship |
| 6 | UCS548 | Data Science |

**References**

[1] “22 essential twitter statistics you need to know in 2023,” The Social Shepherd. [Online]. Available: <https://thesocialshepherd.com/blog/twitter-statistics#:~:text=Twitter%20Has%20396.5%20Million%20Users%20Globally&amp;text=In%20fact%2C%20Elon%20Musk%20recently,what%20others%20are%20tweeting%20about>..

[2] K. Subba Reddy , M. Surya Bhupal rao , Makineedi Raja Babu , P.Naveen Sundar Kumar , A Sophisticated Semantic Analysis Framework using an Intelligent Tweet Data Clustering and Classification Methodologies, Microprocessors and Microsystems (2023)

[3] Ahmed, J., & Ahmed, M. (2023). Classification, detection and sentiment analysis using machine learning over next generation communication platforms. Microprocessors and Microsystems, 98, 104795. <https://doi.org/10.1016/j.micpro.2023.104795>

[4] L. Almuqren and A. I. Cristea, "Predicting STC Customers’ Satisfaction Using Twitter," in IEEE Transactions on Computational Social Systems, vol. 10, no. 1, pp. 204-210, Feb. 2023, doi: 10.1109/TCSS.2021.3135719.

[5] E. Sharma, A. Gaur and S. Singhal, "Twitter Sentiment Analysis of India vs Pakistan T20 World Cup Match Using SVM Classifier," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 16-19, doi: 10.1109/COM-IT-CON54601.2022.9850549.

[6] Swathi, T., Kasiviswanath, N. & Rao, A.A. An optimal deep learning-based LSTM for stock price prediction using twitter sentiment analysis. *Appl Intell* 52, 13675–13688 (2022).

[7] C. Patra, P. S. D and S. Chakraborty, "Analyzing Information Flow of Hashtag Networks during Elections using Sentiment Analysis and Graph Algorithms," 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2022, pp. 1-6, doi: 10.1109/CONECCT55679.2022.9865804.

[8]Demotte, P., Wijegunarathna, K., Meedeniya, D. *et al.* Enhanced sentiment extraction architecture for social media content analysis using capsule networks. *Multimed Tools Appl* (2021).

[9] Gandhi, U.D., Malarvizhi Kumar, P., Chandra Babu, G. et al. in [6] perform  Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM). Wireless Pers Commun (2021).

[10] J. K. Chandra, E. Cambria and A. Nanetti, "One Belt, One Road, One Sentiment? A Hybrid Approach to Gauging Public Opinions on the New Silk Road Initiative," 2020 International Conference on Data Mining Workshops (ICDMW), Sorrento, Italy, 2020, pp. 7-14, doi: 10.1109/ICDMW51313.2020.00011.

[11] L. G. Singh, A. Anil and S. R. Singh, "SHE: Sentiment Hashtag Embedding Through Multitask Learning," in IEEE Transactions on Computational Social Systems, vol. 7, no. 2, pp. 417-424, April 2020, doi: 10.1109/TCSS.2019.2962718.

[12] A. -L. Alten, G. Gadre, S. Kulkarni and C. -S. Wu, "Analyzing Happiness Index Based on Geographical Locations," 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2019, pp. 45-51, doi: 10.1109/ICAIBD.2019.8837010.

[13] Herlawati, R. T. Handayanto, D. Setiyadi and E. Retnoningsih, "Corpus Usage for Sentiment Analysis of a Hashtag Twitter," *2019 Fourth International Conference on Informatics and Computing (ICIC)*, Semarang, Indonesia, 2019, pp. 1-5, doi: 10.1109/ICIC47613.2019.8985772.

[14] V. Rekha, R. Raksha, P. Patil, N. Swaras and G. L. Rajat, "Sentiment Analysis on Indian Government Schemes Using Twitter data," 2019 International Conference on Data Science and Communication (IconDSC), Bangalore, India, 2019, pp. 1-5, doi: 10.1109/IconDSC.2019.8817036.

[15] A. Aslam, U. Qamar, R. A. Khan, P. Saqib, A. Ahmad and A. Qadeer, "Opinion Mining Using Live Twitter Data," 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), New York, NY, USA, 2019, pp. 36-39, doi: 10.1109/CSE/EUC.2019.00016.

[16] G. Yang, H. He and Q. Chen, "Emotion-Semantic-Enhanced Neural Network," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 3, pp. 531-543, March 2019, doi: 10.1109/TASLP.2018.2885775.

[17] S. Dorle and N. Pise, "Political Sentiment Analysis through Social Media," 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2018, pp. 869-873, doi: 10.1109/ICCMC.2018.8487879.

[18] S. Wankhede, R. Patil, S. Sonawane and P. A. Save, "Data Preprocessing for Efficient Sentimental Analysis," *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Coimbatore, India, 2018, pp. 723-726, doi: 10.1109/ICICCT.2018.8473277.

[19] S. Geetha and V. K. Kaliappan, "Tweet Analysis Based on Distinct Opinion of Social Media Users'," 2018 International Conference on Soft-computing and Network Security (ICSNS), Coimbatore, India, 2018, pp. 1-6, doi: 10.1109/ICSNS.2018.8573619.

[20] J. Chen, H. Li, Z. Wu and M. S. Hossain, "Sentiment analysis of the correlation between regular tweets and retweets," 2017 IEEE 16th International Symposium on Network Computing and Applications (NCA), Cambridge, MA, USA, 2017, pp. 1-5, doi: 10.1109/NCA.2017.8171354.

**Formatting Guidelines**

* Project Report Type:  Transparencies and tape bound
* Number of Copies: 1 per Project group (Max pages 15)
* Running text should be justified, figures and tables center aligned, no space before full stop etc.
* Use **passive voice** in text.
* Paper Size (orientation): A4 (portrait)
* Margins: 1” top / bottom / right and 1.5” left
* Font Type: Times New Roman
* Font Size: 16 bold for Section names, 14 bold for headings and 12 for normal text
* Line Spacing: 1.5 throughout
* Page Numbering:  Bottom center of page in the format – Page 1 of N
* All table and figure captions in size 10 sentence case, table captions on top and figure captions below the figure.
* All figures and tables quoted in the text with explanation.
* No figures and equations should be copied. Please use **smartdraw/ visio for figures and Mathtype** for equations.
* References (The listing of references should be typed 2 spaces below the heading “REFERENCES” in alphabetical order in single spacing left – justified.  It should be numbered consecutively (in square [ ] brackets, throughout the text and should be collected together in the reference list at the end of the report. The references should be numbered in the order they are used in the text. The name of the author/authors should be immediately followed by the year and other details). References should not be cited from Blogs, Twitter etc. but should refer to good Journal or Conference papers. Typical examples of the references are given below:

**REFERENCES**

[1] Ariponnammal, S. and Natarajan, S. (1994) ‘Transport Phonomena of SmSel – X Asx’, Pramana – Journal of Physics Vol.42, No.1, pp.421-425.

[2] Anderson T. , Peterson L., Shenker S., Turner J.(2005).Overcoming the Internet impasse through virtualization. IEEE Computer, 38(4):34-41.

[3] W. Zeng, H. Yu, C. Lin. (2013, Dec 19). Multimedia Security Technologies for Digital Rights Management [Online]. Available: http://goo.gl/xQ6doi