International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Real Time Data Twitter Trends Polling Using Rae Model

[1]Jyothi K. S., [2]Dr. Shantala C. P.

Submitted: 13/10/2023 **Revised**: 12/12/2023 **Accepted**: 22/12/2023

Abstract: Social media holds valuable insights into individuals and society, offering a wealth of data to propel research across various domains, like business, finance, health, socio-economic inequality, and gender vulnerability. Within this landscape, Twitter emerges as a prominent platform primarily utilized for emotive expression around specific events. Functioning as a micro-blogging hub, Twitter serves as a conduit for gathering opinions on products, trends, and political discourse. Twitter generates an immense volume of data, contributing significantly to the challenges associated with big data. Among these challenges lies the complexity in classifying tweets, stemming from the intricate and sophisticated language used, rendering existing tools inadequate. Despite extensive efforts dedicated to this issue, there remains a lack of definitive validation aligning online social media trends with conventional survey results. Sentiment analysis emerges as a method aimed at scrutinizing the sentiments, emotions, and viewpoints of diverse individuals regarding various subjects, capable of examining public opinion expressed in tweets related to news, policies, social movements, and influential figures. Sentiment Analysis has leveraged Machine Learning Classifiers, enabling the automation of opinion mining without the need for manual tweet reading. Machine Learning models have consistently demonstrated impressive outcomes across diverse applications. Thus, this study introduces the utilization of the Real-time Advanced Ensemble Learning (RAE) model for live Twitter trend polling based on real-time data. The effectiveness of this approach will be assessed through metrics such as training and validation accuracy, as well as training and validation loss. Expectations suggest that this model will yield notable advancements compared to previous methods.

Keywords**:** *Social Media, Twitter, Sentiment Analysis, Opinion Mining, Machine Learning.*

1. Introduction

The exponential surge in social media users results in the creation of vast volumes of unstructured text encompassing messages, chats, posts, and blogs. Beyond serving as a means of information exchange, social media stands as an incredibly accessible platform for the expression of ideas and opinions, often gaining traction when endorsed by a sizable user base. The widespread usage of these platforms might mirror the collective sentiment individual's harbor toward a particular individual, entity, or location. Online networks like Facebook, Twitter, Flickr, and Instagram have emerged as the favored virtual arenas for communication and social engagement across diverse age groups [4]. Twitter stands as a pivotal online social network (OSN), securing the 42nd position among websites according to Alexa's rankings. Additionally, statistical data indicates a surge in user accounts across various nations; for instance, there are approximately 68.7 million accounts in the United States and 12.35 million accounts in Saudi Arabia [5].

Social media platforms like Twitter produce extensive textual content encompassing political perspectives, ripe for mining to understand public opinion and forecast electoral trends. Within Twitter, users have the capability to "retweet" or share posted tweets within their network, facilitating rapid information dissemination. Consequently, retweets serve as indicators of widespread interest among Twitter users, gauging the popularity of tweets based on their content and the magnitude of retweet activity [8].

Tweets frequently convey diverse opinions on various subjects, holding significance in both business-related choices and political assessments, such as evaluating a candidate's sentiments. Consumers leverage sentiment analysis to investigate products or services before committing to a purchase. For instance, marketers in the Kindle industry utilize this tool to gauge public opinion regarding their company and products, or to assess customer satisfaction. Similarly, election polls and organizations harness sentiment analysis to gather crucial feedback, identifying issues with newly launched products. Brands like Nike and Adidas use this approach for effective brand management as well.

Twitter functions as a rapid and efficient micro-blogging platform where users share concise posts known as tweets. It stands as a highly sought-after application globally, ranking among the successful social media platforms. Twitter offers the opportunity to create free accounts, providing access to an extensive potential

[1]*Research Scholar, Department of Computer Science Engineering, Channabasaveshwara Institute of Technology, Gubbi, Visvesvaraya Technological University, Belagavi, Karnataka.*

[2]*Professor, Department of Computer Science Engineering, Channabasaveshwara Institute of Technology, Gubbi, Visvesvaraya Technological University, Belagavi, Karnataka.*

audience. Particularly in business and marketing endeavors, Twitter emerges as an optimal platform, enabling connections with prominent figures such as celebrities and stars. This association proves enticing for both these personalities and advertisers, fostering engaging purchasing opportunities. Twitter serves as a link between celebrities and their fans, offering a means of direct communication with followers. This platform stands out as one of the finest methods for fostering connections among enthusiasts. Despite its limitation of a short character count, restricting posts to 140 letters each, it allows users to share posts or links on the website without any cost, also providing openness for advertisements. Unlike other social networking sites, there are no issues with excessive personal ads cluttering the platform. The immediacy of Twitter is notable; once a tweet is posted, it promptly reaches the audience following the respective business or individual.

The opinions voiced on Twitter often carry a casual, sincere, and informative tone compared to those gathered through formal surveys. Millions of users freely express their sentiments regarding the brands they engage with. Recognizing and understanding these sentiments holds immense value for companies, enabling them not only to monitor their brand's performance but also to pinpoint specific aspects and timeframes that elicit polarized sentiments. These brands could encompass products, celebrities, events, or political parties [1].

Sentiment analysis serves as a method to dissect the diverse attitudes, emotions, and viewpoints of individuals regarding various subjects. When applied to tweets, this technique delves into public opinion regarding news, policies, social movements, and personalities. The primary aim of tweet sentiment analysis lies in identifying the positive, negative, or neutral sentiment components within the data shared on Twitter [6]. Analyzing sentiments through sentiment analysis proves invaluable for organizations seeking insights into public perceptions of their company and products. This automated process categorizes textual data into positive, negative, or neutral sentiments. Employing machine learning for sentiment analysis on Twitter data aids companies in comprehending the conversations surrounding their brand. Sentiment analysis, also known as Opinion Mining, involves computationally examining the opinions, sentiments, and subjective aspects of text [10].

Tweets originate from varied sources, often reflecting dynamic viewpoints and covering multifaceted information in an unstructured, unrefined format. Nonetheless, analyzing the expressed sentiments poses considerable challenges. Issues pertaining to tonality, polarity, lexicon, and tweet grammar complicate this task, given their highly informal and quasi-grammatical nature. Understanding their context becomes challenging, especially considering the prevalent use of slang, acronyms, and internet-specific vocabulary. Categorizing these terms based on polarity proves intricate for natural language processors involved in sentiment analysis [14].

The thoughts, opinions, and sentiments of the general public significantly influence election outcomes. Currently, assessing presidential approval polls involves time-consuming and costly telephone sampling of small groups. An automated method utilizing readily accessible social data would prove immensely beneficial in reducing both time and costs involved in conducting these polls. Moreover, technological advancements highlight the reliability and cost-effectiveness of online platforms in predicting results, showcasing their increasing significance in this domain.

Recently, there's been a recognition that traditional polling methods might not provide the precision and accuracy desired for predictions. Consequently, scientists and researchers have shifted their focus towards examining and analyzing web data, including blogs and social network activity, as an alternative approach to predict election results, aiming for enhanced accuracy. Additionally, conventional polling techniques tend to be costly, contrasting with the accessibility of online information that can be acquired without financial expenditure. Despite extensive research, several challenges persist in this domain.

As social media platforms continue to proliferate, more individuals are voicing their sentiments and opinions. Categorizing these sentiments proves invaluable for those seeking to leverage these opinions to enhance their products, services, and more. Sentiment Analysis stands as the method of delving into textual content, mining it to discern and determine the prevailing sentiment or opinion within the text. As the populace increasingly engages with social media, examining the online information has become instrumental in understanding shifts in people's perceptions, behaviors, and psychology. Consequently, leveraging Twitter data for sentiment analysis has emerged as a prevalent practice. The heightened interest in analyzing social media has sparked greater attention toward Natural Language Processing (NLP) and Artificial Intelligence (AI) technologies focused on text analysis [3].

Machine learning is essential in accurately classifying customer sentiment due to the potential for human error among experts in sentiment analysis. Through the utilization of natural language processing and machine learning techniques, subjective text awareness can be extracted and categorized based on polarity—negative, neutral, or positive [2].

The advancement in Machine and Deep Learning has significantly enhanced the capability to forecast sentiment in textual data. These algorithms have enabled remarkably accurate sentiment predictions. Machine learning facilitates learning new tasks without explicit training or programming. Moreover, sentiment analysis models extend beyond predicting sentiment alone, encompassing other subjective information within text. Various machine learning and deep learning algorithms can be employed to predict emotions and sentiments.

Accurately assigning sentiment classes on a large scale necessitates thorough analysis. This involves the utilization of precise NLP techniques and machine learning (ML) models tailored for text classification. Twitter offers a vast and comprehensive dataset for analysis, requiring efficient methods to automatically label its inherently noisy text data. Despite numerous studies on Twitter sentiment classification in the past, their effectiveness has been limited. Therefore, a real-time solution addressing these challenges involves polling real-time Twitter trends data using the RAE model.

The subsequent sections are structured as follows: Section II provides an overview of the literature surveyed. Section III showcases the real-time polling of Twitter trends data using the RAE model. Following that, Section IV assesses the results derived from the proposed approach. The work culminates with a conclusion in Section V.

2. Literature Survey

Abu Nowhash Chowdhury, Shawon Guha, Nurul Amin, Shahidul Islam Khan, et al. [7] outline the methodology in "Exploiting Diverse Contextual Features through Transformers for Detecting Informative Tweets." Their approach involves harnessing an ensemble of cutting-edge transformer models to capture the diverse contextual dimensions present in tweets. This method utilizes a combination of BERT (Bidirectional Encoder Representations from Transformers), CTBERT, BERTweet, RoBERTa (Robustly Optimized BERT Pre-training Approach), and XLM-RoBERTa (Cross-lingual Language Model) models. Subsequently, a pooling operation is applied to the extracted embedding features to convert them into document embedding vectors. Following this, a feed-forward neural architecture incorporating a linear activation function is employed for the classification task. The final predictions are generated using a majority voting-driven ensemble technique. Experimental trials conducted on the WNUT-2020 (Workshop on Noisy User-generated Text) COVID-19 English Tweet dataset demonstrated the superior efficacy of our method compared to other state-of-the-art approaches.

D. Sunitha, Raj Kumar Patra, N.V. Babu, A. Suresh, Suresh Chand Gupta, et al. [9], discuss Twitter sentiment analysis focusing on COVID-19 in India and European countries, employing an ensemble-based deep learning model. Their model aims to analyze real-time tweets pertaining to the coronavirus. Initially, approximately 3100 tweets from individuals in India and Europe were gathered spanning the period from 23rd March 2020 to 1st November 2021. Subsequently, data preprocessing and exploratory analysis were conducted to gain a comprehensive understanding of the collected data. Additionally, feature extraction involves employing various methods such as Term Frequency-Inverse Document Frequency (TF-IDF), GloVe, pre-trained Word2Vec, and fast text embeddings. These extracted feature vectors are then inputted into the ensemble classifier comprising Gated Recurrent Unit (GRU) and Capsule Neural Network (CapsNet) to classify user sentiments into categories like anger, sadness, joy, and fear. The experimental results demonstrated that the proposed model achieved notably high prediction accuracy in effectively classifying sentiments of both Indian and European individuals.

Piyush Vyas, Martin Reisslein, Bhaskar Prasad Rimal, Gitika Vyas, Ganga Prasad Basyal, Prathamesh Muzumdar, et al. [11], introduce an innovative framework titled "Automated Classification of Societal Sentiments on Twitter With Machine Learning." This framework stands out as it combines a lexicon-based approach for analyzing and labeling tweet sentiments with supervised ML techniques for tweet classification in a hybrid manner. The hybrid framework underwent evaluation across multiple metrics including precision, accuracy, recall, and F1 score. Findings revealed that the predominant sentiments observed were largely positive (38.5%) or neutral (34.7%). Notably, with an 83% accuracy rate, the long short-term memory (LSTM) neural network emerged as the preferred ML technique within this framework. These evaluation outcomes underscore the potential of this hybrid framework to autonomously categorize extensive volumes of tweets, including those concerning COVID-19, based on societal sentiments.

Umit Demirbaga and colleagues [12] introduce HTwitt, a Hadoop-based platform tailored for analyzing and visualizing streaming Twitter data. HTwitt is structured atop the Hadoop ecosystem, integrating a MapReduce algorithm and a suite of machine learning techniques within a big data analytics framework. This endeavor addresses challenges by employing various algorithms alongside a Naive Bayes classifier to ensure both reliability and highly precise recommendations within virtualization and cloud environments. The primary contribution lies in presenting a framework aimed at constructing landslide early warning systems, achieved by identifying valuable tweets and visualizing them alongside processed information. The experimental results illustrate the degrees of overfitting in the model's training phase across diverse sizes of real-world datasets in the machine learning stages.

Payal Khurana Batra, Aditi Saxena, Shruti, Chaitanya Goel, et al. [15], delve into "Election Result Prediction Using Twitter Sentiment Analysis," foreseeing outcomes grounded in news, discussions, and online platforms. Contemporary trends emphasize the significance of online sites like Facebook, Twitter, and WhatsApp. Accessible to anyone with an internet connection, these social platforms have gained widespread usage in today's context. The insights, news, and discussions found on these platforms serve as valuable predictors for future outcomes. They provide a virtual space for individuals to express their opinions. Notably, Twitter stands out as a prominent platform, particularly during significant national events. News quickly transforms into trends on this platform, where people openly express their opinions, criticisms, and campaign sentiments toward political figures or parties. This widespread engagement becomes instrumental in anticipating electoral outcomes even before the release of exit poll results. Consequently, this study focuses on analyzing tweets sourced from Twitter, employing sentiment analysis to predict election results based on this data.

Anam Yousaf, Muhammad Umer, Saima Sadiq, Saleem Ullah, Seyedali Mirjalili, Vaibhav Rupapara, Michele Nappi, et al. [16], explore "Emotion Recognition by Textual Tweets Classification Using Voting Classifier (LR-SGD)." Leveraging Machine Learning models allows for automated opinion mining without the need for manual reading of tweets. Their findings hold the potential to guide governments and businesses in shaping policies, products, and events. The study implements seven Machine Learning models to recognize emotions, categorizing tweets into happy or unhappy sentiments. An extensive comparative analysis of performance revealed that the proposed voting classifier (LR-SGD) using TF-IDF generated the most optimal outcome, achieving an accuracy of 79% and an F1 score of 81%. Additionally, validating the stability of this approach on two additional datasets, one binary and the other multi-class, yielded robust and consistent results.

Yogesh Chandra, Antoreep Jana, et al. [18], detail "Sentiment Analysis using Machine Learning and Deep Learning," utilizing diverse methodologies for sentiment analysis. Machine Learning Classifiers are employed for sentiment analysis, encompassing polarity-based sentiment analysis alongside Deep Learning Models to categorize users' tweets as either 'positive' or 'negative' in sentiment. The incorporation of various model architectures aims to accommodate the diversity of opinions and thoughts prevalent on social media platforms. These classification models possess the potential for implementation in categorizing real-time tweets on Twitter across various topics.

Manoj Sethi, Sarthak Pandey, Prashant Trar, Prateek Soni, et al. [19], delve into "Sentiment Identification in COVID-19 Specific Tweets." Their aim is to offer a domain-specific method to comprehend the sentiments expressed globally concerning this particular situation. To achieve this, tweets specific to the coronavirus are collected from the Twitter platform. Following the collection of tweets, they undergo labeling, paving the way for the development of an efficient model designed to discern the true sentiment conveyed in COVID-19-related tweets. Comprehensive evaluations are conducted in both binary and multi-class contexts, utilizing n-gram feature sets and encompassing cross-dataset assessments of various machine learning techniques to refine the model. Our experimental findings highlight the efficacy of the proposed model in accurately capturing people's perceptions of COVID-19, showcasing a high level of accuracy.

3. Real Time Data Twitter Trends Polling

The following section demonstrates the utilization of the Real-time Advanced Ensemble Learning (RAE) model for polling Twitter trends via real-time data. Figure 1 illustrates the block diagram detailing this process.

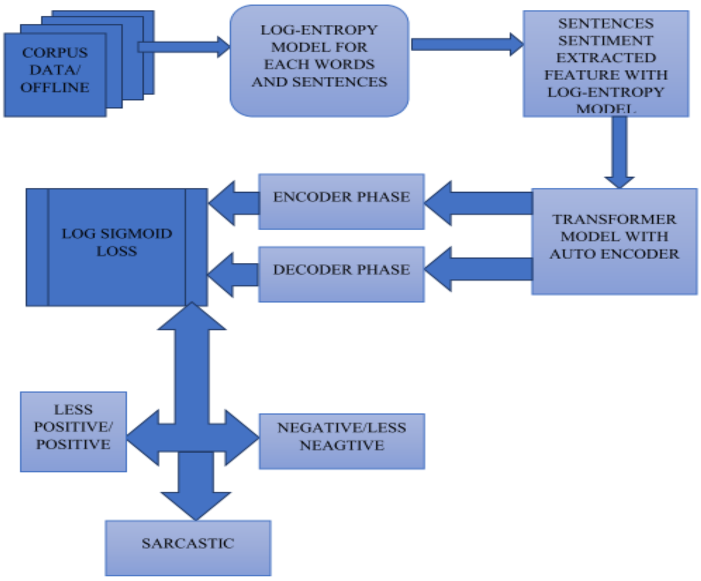

Fig. 1: Block diagram of real time data witter trends polling using RAE model

The steps of presented approach are as follows: i) Log in to twitter API (Application Programming Interface) with developers account for access. ii) Define a function to inculcate all the recent trends based on the search feature. iii) Apply the different hash tags and its polling feature to extract from the recent trends. iv) Initiate a scenario of different tweets for list of hash tags. v) Initiate a calling functionality for the hash tags and its conversion into trends. vi)Polling using twython. vii) Removing stop words. viii) Polling using hash tags. ix) Using improvised LDA weights and x) Machine learning model.

Upon viewing Twitter content like embedded Tweets, buttons, or timelines integrated into external websites via Twitter for Websites, Twitter may acquire data, including the web page you accessed, your IP address, browser type, operating system details, and cookie information.

The Twitter API enables you to programmatically execute numerous actions akin to those achievable through the Twitter app or website. These actions encompass searching for Tweets, following users, reading your timeline, posting tweets and direct messages, and more. However, certain aspects of the Twitter user experience, like polls, are currently inaccessible through the API (as of this writing). Leveraging the API allows you to gather data from tweets and create automated agents for posting on Twitter.

Using polling involves passing a function, a step (time interval between calls), and a timeout. By default, your function will be repeatedly called until it returns a truthy value or the cumulative execution time surpasses the specified timeout. Twitter's API functions over HTTP (Hyper Text Transfer Protocol), delivering JSON objects in response. In theory, you could utilize the requests library (or any other HTTP client) to interact with the API, making requests and receiving responses. Nevertheless, the Twitter API employs a rather intricate authentication mechanism known as Oauth. It mandates the creation of cryptographic signatures for requests to uphold their security. Implementing this process from scratch can be intricate and isn't necessarily worth the effort. Hence, most programmers leveraging the Twitter API opt for third-party libraries. These tools encapsulate and simplify the intricacies of Oauth authentication, alleviating concerns for programmers. Additionally, these libraries offer specific abstractions for API calls, making API utilization more straightforward—you can simply invoke methods with parameters instead of manually constructing URLs within your code.

Several libraries offer access to the Twitter API. We'll utilize one called Twython for this purpose, which can be installed using pip.

Typically, Twython corresponds to each "endpoint" within the Twitter REST API with a method. These methods often mirror or directly match the URLs found in the REST API. In the Twython API documentation, available methods and their mapping to different parts of the Twitter API are clearly listed.

To gain a better understanding, let's delve deeper into the search functionality. Twython's `.search()` method aligns with various parameters mirroring the query string parameters found in the REST API's search tweets endpoint, detailed in the documentation here. Essentially, every parameter applicable in the query string of a REST

API call can similarly be included as a named parameter within Twython's `search()` method call.

The call to `search()` involves certain parameters: `q` (defining the search query), `result_type` (adjustable between popular, recent, or mixed for desired result returns), and `counts` (determining the number of tweets in the response, capped at 100). Upon reviewing the documentation, an intriguing parameter, `geo code`, seems available for exploration. It enables the search to retrieve tweets solely within a specified radius of a particular latitude/longitude. Employing diverse hashtags aids in polling to engage a wider user base.

The corpus is specifically curated for sentiment analysis and opinion mining. Through linguistic analysis of this collected corpus, we aim to elucidate discovered phenomena. Leveraging this corpus, we construct a sentiment classifier capable of discerning between positive, negative, and neutral sentiments. These corpora can encompass texts within a single language (monolingual corpus) or incorporate text data across multiple languages (multilingual corpus).

To enhance the utility of corpora for linguistic research, they commonly undergo a process called annotation. For instance, part-of-speech tagging (POS-tagging) involves adding information regarding each word's part of speech (verb, noun, adjective, etc.) to the corpus in the form of tags. Similarly, indicating the lemma (base) form of each word is another example of annotation.

In the field of linguistics, a corpus or text corpus refers to a language resource containing an extensive and organized compilation of texts. In corpus linguistics, they serve for statistical analysis and hypothesis testing, investigating occurrences or confirming linguistic rules within a particular language domain. In search technology, a corpus represents the aggregate of documents being searched. It can comprise texts in a singular language (monolingual corpus) or encompass text data across multiple languages (multilingual corpus).

To enhance the utility of corpora for linguistic research, they commonly undergo annotation, a process involving various tasks. One such task is part-of-speech tagging (POS-tagging), which supplements the corpus with information regarding each word's part of speech (verb, noun, adjective, etc.) in the form of tags. Another annotation example involves specifying the lemma (base) form of each word. To bridge the language barrier between researchers and the corpus's language, interlinear glossing is employed to provide bilingual annotation. Guaranteeing comprehensive and consistent annotation across the entire corpus poses a challenge, resulting in smaller corpora typically comprising one to three million words. Additional levels of linguistic structured analysis, involving various annotations, are also viable. for morphology, semantics and pragmatics.

Eliminating stop words stands as a prevalent preprocessing step in various NLP applications. The concept revolves around excluding words that frequently appear across all documents in the corpus. Removing stop words from multi-word queries aims to enhance search performance.

The log-entropy model facilitates the transformation of a basic Bag of Words (BoW) represented corpus into log entropy space. It operates using the Log Entropy Model, generating an entropy-weighted logarithmic term frequency representation. Instances of this class enable the conversion between a word-document co-occurrence matrix (int) into a matrix weighted either locally or globally (positive floats).

This process involves log entropy normalization, which may include an additional step of normalizing the resulting documents to unit length. The subsequent formulas outline how to compute the log entropy weights for term i within a document. *j*:

$$local_{weight_{i,j}} = \log(frequency + 1)$$

$$P_{i,j} = \frac{frequncy_{i,j}}{\sum_j frequency_{i,j}}$$

$$global_{weight_i} = 1 + \frac{\sum_j P_{i,j} * \log(P_{i,j})}{\log(number_{of_{documents}} + 1)}$$

$$final_{weight_{i,j}} = local_{weight_{i,j}} * global\_weight_i$$

Parameters: corpus (iterable of iterable of (int, int)) – Input corpus in BoW format.

normalize (bool, optional) – If True, the resulted log entropy weighted vector will be normalized to length of 1, If False - do nothing.

The encoder comprises encoding layers that sequentially process the input, while the decoder includes decoding layers that similarly handle the output from the encoder. Each encoder layer aims to create encodings that encapsulate the interrelated information within the inputs, forwarding these encodings to subsequent encoder layers. Conversely, every decoder layer leverages the collective contextual information from all encodings to generate an output sequence. Both encoder and decoder layers employ an attention mechanism to accomplish this task.

Within the input, attention assesses the significance of each segment in relation to others, using this insight to generate the output. Every decoder layer incorporates an extra attention mechanism, leveraging information from prior decoder outputs before the layer accesses details from the encodings. The fig. 2 shows the flow diagram of Transformer Auto encoder.

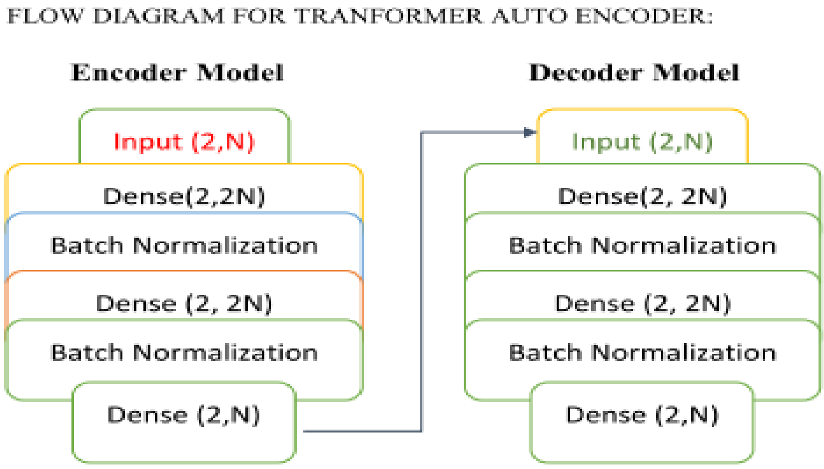

Fig. 2: Flow diagram of Transformer Auto encoder

The Transformer encoder processes an input sequence following word embedding and positional encoding stages.

Batch normalization standardizes network inputs, applied to either preceding layer activations or direct inputs. It expedites training, sometimes halving epochs or better, and offers regularization, diminishing generalization error. Normalization involves transforming data to a mean of zero and standard deviation of one. Here, with our batch input from layer h, the initial step involves computing the mean of this hidden activation.

$$\mu = \frac{1}{m}\sum h_i \quad (1)$$

where, m is the number of neurons at layer h. Once we have meant at our end, the next step is to calculate the standard deviation of the hidden activations.

$$\sigma = [\frac{1}{m}\sum (h_i - \mu^2]^{\frac{1}{2}} \quad (2)$$

Further, as we have the mean and the standard deviation ready. We will normalize the hidden activations using these values. For this, we will subtract the mean from each input and divide the whole value with the sum of standard deviation and the smoothing term (ε). The smoothing term(ε) assures numerical stability within the operation by stopping a division by a zero value.

$$h(norm) = \frac{(h_i - \mu)}{\sigma + \varepsilon} \quad (3)$$

These two are learnable parameters, during the training neural network ensures the optimal values of γ and β are used. That will enable the accurate normalization of each batch. The algorithm description is as follows:

ALGORITHM:

Input: Sentences Extracted Features (x_n), w_i weights for the Each sentences

Output: Sentences sentiments (y_n)

Let m_i be the expected weights for each sentiments observed,

E be the expected function

For I in range (0, length (dataset)):

For jin range of (i[th] column):

$y_n(0) \leftarrow 0$

$$y(n+1) \leftarrow (log_{10} e^{w^i * x^i / (1 - E(x_i))})^{-1/2} * \big(y(n)\big) + E(\frac{w_i}{x_i})$$

let ε be the loss estimated with the flow model of the design indicating the overall sigmoid loss calculated as mentioned below:

$$E\left(\varepsilon_i\big(y(n+1)\big)\right) = \mu \sum_{i=0}^{N} x_i\, W_i - \sigma y(n)/(1 - e^{xi}$$

Therefore, employing the aforementioned formulations, we've effectively estimated the overall loss of the model in sentiment analysis.

Sigmoid Cross-Entropy Loss, also known as Binary Cross-Entropy Loss, combines Sigmoid activation with Cross-Entropy loss. It serves as the default loss function for binary classification tasks. To classify data into two classes, it necessitates a single output layer with an output range of 0 to 1, hence employing the sigmoid function. For multi-class classification, Linear Discriminant Analysis (LDA) stands as a well-received technique, adept at learning the most distinguishing features.

Advanced Ensemble Learning Techniques in Real-Time: The array of advanced learning techniques includes Stacking, Blending, Bagging, and Boosting.

4. Result Analysis

In this segment, we implement real-time Twitter trend polling using the RAE model. The ensuing analysis of the approach's results is conducted, evaluating its performance based on validation loss, training loss, validation accuracy, and training accuracy. The accuracy metric is defined as follows:

Accuracy: It's delineated as the quotient of correct predictions to the total predictions made, expressed as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) represent the instances utilized in polling predictions. Figure 3 demonstrates the polling for various hashtags.

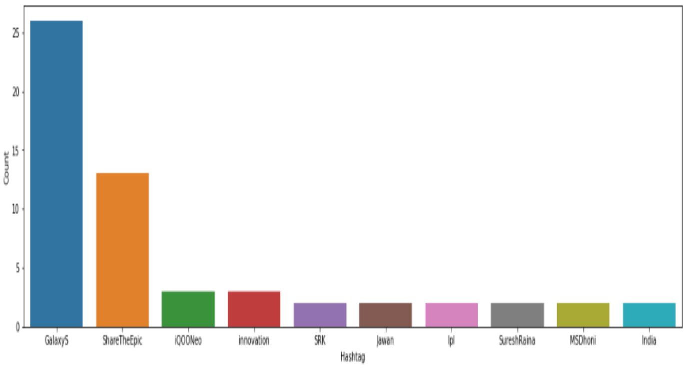

Fig. 3: polling of Different Hash tags

The evaluation of the presented approach hinges on the validation loss and validation accuracy. Learning curves depict how the model performs on both the training and validation sets, varying according to different sample sizes from the training dataset. Specifically, these curves illustrate training and validation scores on the y-axis against different sample sizes of the training dataset on the x-axis.

The training loss reflects the model's fitting to the training data, whereas the validation loss gauges its adaptation to new data. It's a crucial metric to evaluate model performance on the validation set, a dedicated portion of the dataset used for this purpose. The loss value represents the aggregate errors across examples in both training and validation sets, serving as a measure of the model's behavior after each optimization iteration.

The accuracy metric serves as a comprehensible measure of the algorithm's performance. In machine learning, it's common for the training accuracy to slightly exceed the testing accuracy as the model utilizes the training data for predictions. Figure 4 illustrates the accuracy for both training and validation datasets.

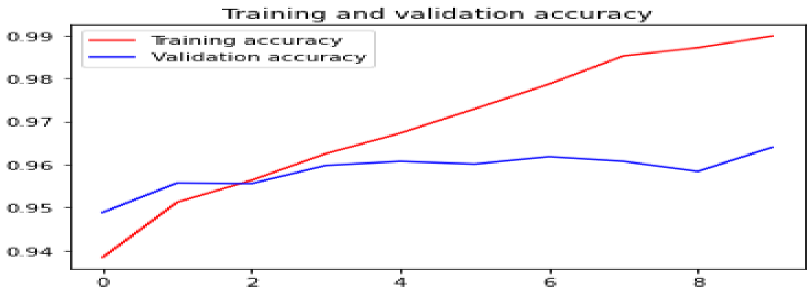

Fig. 4: Training and Validation Accuracy curves

Figure 4 illustrates that the training accuracy surpasses the validation accuracy, with the former almost reaching 1 while the latter is at 0.96. Additionally, Figure 5 depicts the training and validation loss.

Fig. 5: The validation and training loss

From fig. 5, it is observed that the minimum validation loss is 0.11 and the training loss is 0.03. The Fig.6 shows the polling results and accuracy.

```
Can you imagine how this idea's gonna change the education system!
0 : 0.9988957643508911


Brilliant! I can't wait to hear the news about this change!
0 : 0.9972726702690125


Chill out guys, nothing's gonna change, we have to study hard to succeed
0 : 0.9996966123580933
```

Fig. 6: Polling Results and Accuracy

Thus, the utilization of the RAE model in polling real-time Twitter trends resulted in highly accurate and effective predictions across various polling outcomes on Twitter.

5. Conclusion

This project showcases the utilization of an RAE model for real-time Twitter trend polling. To begin, access the Twitter API by logging in with a developer's account. Then, establish a function to gather the latest trends using the search feature. Following this, employ various hashtags and their polling attributes to extract up-to-date trends. Lastly, simulate a scenario involving multiple tweets corresponding to a list of hashtags. The process begins by activating a function for converting hashtags into trends, utilizing Twython for polling purposes. Enhancing the search engine's efficiency involves removing stop words, improving the polling process across various hashtags. To refine the polling results, an RAE model is employed to enhance LDA weights. The performance evaluation encompasses metrics such as validation loss, training loss, validation accuracy, and training accuracy. Notably, the training loss surpasses the validation loss, while training accuracy outperforms validation accuracy. Additionally, the polling accuracy for diverse tweets is notably precise.

References

[1] Suraj Kumar, Sonu Kumar, Shivam Bharatiya, Shanvi Gautam, Prof. Mr.Shakun Garg, "Real Time Twitter Sentimental Analysis", International

Journal of Creative Research Thoughts (IJCRT), 2023 IJCRT, Volume 11, Issue 3 March 2023

[2] Niklas Braig, Alina Benz, Soeren Voth, Johannes Breitenbach, Ricardo Buettner, “Machine Learning Techniques for Sentiment Analysis of COVID-19-Related Twitter Data”, IEEE Access (Volume: 11), 2023, doi: DOI: 10.1109/ACCESS.2023.3242234

[3] Yuxing Qi, Zahratu Shabrina, “Sentiment analysis using Twitter data: a comparative application of lexicon- and machine- learning- based approach”, Social Network Analysis and Mining (2023), doi:10.1007/s13278-023-01030-x

[4] Belal Abdullah Hezam Murshed; Jemal Abawajy; Suresha Mallappa; Mufeed Ahmed Naji Saif; Hasib Daowd Esmail Al-Ariki, “DEA-RNN: A Hybrid Deep Learning Approach for Cyberbullying Detection in Twitter Social Media Platform”, IEEE Access (Volume: 10),

[5] Shikah J. Alsunaidi, Rawan Talal Alraddadi, Hamoud Aljamaan, “Twitter Spam Accounts Detection Using Machine Learning Models”, 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN), DOI: 10.1109/CICN56167.2022.10008339

[6] Abu Nowhash Chowdhury, Shawon Guha; Nurul Amin, Shahidul Islam Khan, “Exploiting Diverse Contextual Features through Transformers for Detecting Informative Tweets”, 2022 International Conference on Innovations in Science, Engineering and Technology (ICISET), DOI: 10.1109/ICISET54810.2022.9775891

[7] Maryam Mahdikhani, “Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of Covid-19 pandemic”, International Journal of Information Management Data Insights 2 (2022) 100053,

[8] D. Sunitha, Raj Kumar Patra, N.V. Babu, A. Suresh, Suresh Chand Gupta, “Twitter sentiment analysis using ensemble based deep learning model towards COVID-19 in India and European countries”, Pattern Recognition Letters 158 (2022) 164–170, doi:10.1016/j.patrec.2022.04.027

[9] Sakirin Tam, Rachid Ben Said, Ö. Özgür Tanriöver, “A ConvBiLSTM Deep Learning Model-Based Approach for Twitter Sentiment Classification”, IEEE Access (Volume: 9), 2021,Electronic ISSN: 2169-3536,

[10] Piyush Vyas, Martin Reisslein, Bhaskar Prasad Rimal, Gitika Vyas, Ganga Prasad Basyal; Prathamesh Muzumdar, “Automated Classification of Societal Sentiments on Twitter With Machine Learning”, IEEE Transactions on Technology and Society (Volume: 3, Issue: 2, June 2021), DOI: 10.1109/TTS.2021.3108963

[11] Umit Demirbaga, “HTwitt: a hadoop-based platform for analysis and visualization of streaming Twitter data”, Neural Computing and Applications, 2021, Springer, doi:10.1007/s00521-021-06046-y

[12] Mohit Dagar, Abhishek Kajal, Pardeep Bhatia, “Twitter Sentiment Analysis using Supervised Machine Learning Techniques”, 2021 5th International Conference on Information Systems and Computer Networks (ISCON),

[13] Arnab Roy, Muneendra Ojha “Twitter sentiment analysis using deep learning models”, 2020 IEEE 17th India Council International Conference (INDICON), 978-1-7281-6916-3/20, 2020 IEEE, DOI: 10.1109/INDICON49873.2020.9342279

[14] Payal Khurana Batra, Aditi Saxena, Shruti, Chaitanya Goel, “Election Result Prediction Using Twitter Sentiments Analysis”, 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC),

[15] Anam Yousaf, Muhammad Umer, Saima Sadiq, Saleem Ullah, Seyedali Mirjalili, Vaibhav Rupapara, Michele Nappi, “Emotion Recognition by Textual Tweets Classification Using Voting Classifier (LR-SGD)”, IEEE Access (Volume: 9), 2020, doi: DOI: 10.1109/ACCESS.2020.3047831

[16] Sheresh Zahoor; Rajesh Rohilla, “Twitter Sentiment Analysis Using Machine Learning Algorithms: A Case Study”, 2020 International Conference on Advances in Computing, Communication & Materials (ICACCM),

[17] [18] Yogesh Chandra; Antoreep Jana, “Sentiment Analysis using Machine Learning and Deep Learning”, 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom),

[18] Manoj Sethi, Sarthak Pandey, Prashant Trar; Prateek Soni, “Sentiment Identification in COVID-19 Specific Tweets”, 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC),

[19] Tulika Saha; Sriparna Saha; Pushpak Bhattacharyya, “Tweet Act Classification : A Deep Learning based Classifier for Recognizing Speech Acts in Twitter”, 2019 International Joint Conference on Neural Networks (IJCNN), DOI: 10.1109/IJCNN.2019.8851805.