IJITCE

ISSN 2347- 3657

International Journal of
Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

Dynamic Sentiment Classifier Using Recurrent Neural Networks to Classify Sentiment in Real-Time Across Multiple Languages for Global Markets

Nadia Shaikh[1], Sudi Shylaja[2], Aljapur Sahasra[3], S. Bavankumar[4]

[1,2,3]UG Scholar, Department of Computer Science and Engineering, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100

[4]Assistant Professor, Department of Computer Science and Engineering, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100

[4]sbavankumar55@gmail.com

Abstract:

The surge of globalization has transformed the market landscape by connecting organizations and people on an international level. A study by the Indian market research firm IMRB indicated that over 70% of Indian consumers rely on online reviews before making purchasing decisions. Furthermore, the exponential growth of the digital landscape, with approximately 600 million Internet users in India, resulted in vast amounts of unstructured data needing analysis. To thrive in this environment, companies must capture customer sentiments on products, services, and trends. Catering to a global audience requires understanding diverse languages and cultural nuances. With the rapid increase in digital data from reviews, surveys, and social media, manual analysis is infeasible. Traditionally, sentiment analysis was conducted manually by experts reviewing customer feedback and social media posts, which was time-consuming and often subjective. To overcome the disadvantages of the former, this project has been introduced. This project proposes a Dynamic Sentiment Classifier utilizing Recurrent Neural Networks, such as LSTM and GRU, optimized to process real-time data across multiple languages. The RNN model captures temporal and contextual relationships, essential for accurately classifying sentiments. Additionally, by leveraging Natural Language Processing (NLP) techniques and multilingual embeddings, the classifier ensures subtle nuances are preserved, making sentiment classification reliable and culturally adaptable. This model equips businesses with actionable insights in real-time, enabling them to fine-tune their strategies and engage effectively with a global audience.

Keywords: Sentiment Classifier, Recurrent Neural Networks, LSTM, GRU, Natural Language Processing

1. INTRODUCTION

In today's interconnected world, businesses operate in an increasingly global and competitive marketplace, where understanding consumer sentiment is a fundamental requirement for success. Consumer sentiment, as expressed through online reviews, social media discussions, and other digital interactions, serves as a treasure trove of insights for businesses. These insights can reveal customer preferences, highlight emerging market trends, and provide critical feedback on products and services. The importance of sentiment analysis has grown exponentially with the proliferation of digital platforms and the reliance of consumers on these platforms for their decision-making processes. A study by the Indian Market Research Bureau (IMRB) found that over 70% of Indian consumers consult online reviews before making purchasing decisions. This statistic underscores a significant shift in consumer behavior, where opinions expressed online heavily influence purchase outcomes. Businesses that can effectively interpret and respond to this feedback gain a competitive edge, enhancing customer engagement, improving their product offerings, and strengthening their brand reputation. However, the sheer volume of data generated by these digital interactions presents a significant challenge. Traditional sentiment analysis methods, such as manual reviews or basic keyword-based techniques, struggle to process this data efficiently and accurately. These approaches often miss the subtleties of language, including context, tone, and cultural nuances, and are ill-equipped to handle multilingual data, which is increasingly common in global markets.

The Dynamic Sentiment Classifier aims to address these challenges by leveraging cutting-edge technologies to perform real-time sentiment analysis at scale. Utilizing Recurrent Neural Networks (RNNs), the system is designed to process complex and unstructured textual data with remarkable precision. Advanced natural language processing (NLP) techniques are integrated into the model to enable it to analyze sentiment across multiple languages, a feature critical for businesses operating in diverse and multilingual regions. Key components of the system include Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), which are specialized architectures within RNNs that excel in handling sequential data. These architectures allow the system to capture long-term dependencies and contextual information, making it adept at understanding the intricacies of consumer opinions expressed in text. By offering real-time sentiment classification, the Dynamic Sentiment Classifier provides businesses with timely insights that can inform their decision-making processes. Whether it's tailoring marketing strategies, refining customer service approaches, or monitoring brand reputation, the system empowers organizations to act proactively. This ability to respond swiftly and accurately to evolving consumer sentiments positions businesses to thrive in today's fast-paced and digitally driven environment. Moreover, the capability to analyze sentiment across multiple languages enhances inclusivity and allows companies to connect with a broader audience, further solidifying their presence in the global market.

Finally, the challenge of scalability cannot be ignored. As businesses grow and expand into new markets, the volume of data they must analyze increases exponentially. Manual methods or systems that lack robust machine learning capabilities struggle to scale effectively, leading to bottlenecks and reduced efficiency. This scalability issue becomes even more critical in scenarios where businesses need to monitor sentiment across multiple regions or product categories simultaneously. Without a scalable and efficient solution, companies face significant obstacles in harnessing the full potential of consumer sentiment data. In the digital age, where vast amounts of textual data are generated every second, the ability to analyse sentiments in real-time is critical. Social media platforms like Twitter and Facebook host billions of multilingual posts daily. Businesses require immediate insights to adapt marketing strategies, address customer complaints, and monitor brand reputation. Politicians and policymakers need real-time feedback on public sentiment to make informed decisions. Real-time sentiment analysis helps e-commerce platforms optimize product recommendations and pricing. Additionally, it is indispensable for crisis management, allowing organizations to detect and respond to negative trends quickly, ensuring consumer trust and satisfaction

2. LITERARURE SURVEY

Indhraom Prabha M. and G. Umarani Srikanth (2019) [1] conducted a comprehensive survey on sentiment analysis utilizing deep learning techniques. Their study delved into various deep learning models, highlighting their effectiveness in capturing complex patterns in textual data. They emphasized the superiority of these models over traditional methods, particularly in handling large-scale data and understanding contextual nuances. The survey also discussed challenges such as computational requirements and the need for extensive labeled data, providing insights into potential areas for future research.

M. Nivaashini, R.S. Soundariya, and P. Thangaraj (2018) [3] performed a comparative analysis of machine learning approaches for Twitter sentiment analysis. Their research evaluated various classifiers, including Naive Bayes, Support Vector Machines, and Decision Trees, assessing their performance in terms of accuracy and computational efficiency. The study found that certain classifiers outperformed others in specific scenarios, suggesting that the choice of algorithm should be context-dependent. They also highlighted the challenges posed by the informal and concise nature of tweets, which can affect sentiment classification accuracy.

Mariam Biltawi et al. (2016) [4] conducted a survey focusing on sentiment classification techniques for the Arabic language. They explored various methods, including machine learning and lexicon-based approaches, discussing their applicability and effectiveness in processing Arabic text. The study highlighted the unique challenges associated with Arabic sentiment analysis, such as dialectal variations and the complexity of morphological structures. The authors emphasized the need for developing more sophisticated tools and resources to improve sentiment analysis in the Arabic context.

Muhammad Rehan and colleagues (2021) [5] proposed the Employees Reviews Classification and Evaluation (ERCE) model using supervised machine learning approaches. Their research aimed to analyze employee reviews to extract valuable insights regarding workplace satisfaction and organizational culture. The study employed various classifiers and evaluated their performance, concluding that the ERCE model could effectively categorize reviews and assist in organizational assessments. They also discussed the implications of their findings for human resource management and organizational development.

Ronglei Hu et al. (2018) [6] provided a comprehensive review of text sentiment analysis methodologies. Their work encompassed both traditional machine learning techniques and emerging deep learning models, analyzing their strengths and limitations. The review highlighted the evolution of sentiment analysis approaches, noting the shift towards deep learning due to its ability to model complex linguistic patterns. The authors also identified ongoing challenges, such as the need for large annotated datasets and the handling of sarcasm and irony in text.

B. Schuller, J.-G. Ganascia, and Laurence Devillers (2016) [7] discussed ethical considerations in multimodal sentiment analysis, particularly concerning data collection, annotation, and exploitation. They emphasized the importance of addressing privacy concerns and obtaining informed consent from data subjects. The paper also highlighted the potential biases that can arise during data annotation and the need for transparency in the development and deployment of sentiment analysis systems. The authors called for the establishment of ethical guidelines to govern research and applications in this field.

Basant Agarwal and Namita Mittal (2014) [8] explored various machine learning approaches for sentiment analysis, comparing their effectiveness in different contexts. Their study examined algorithms such as Support Vector Machines, Naive Bayes, and Maximum Entropy, evaluating their performance on diverse datasets. The authors discussed feature selection techniques and the impact of data preprocessing on model accuracy. They concluded that no single algorithm is universally superior, and the choice of method should consider the specific characteristics of the data and the application domain.

Felix Greaves et al. (2013) [9] investigated the use of sentiment analysis to capture patient experiences from free-text comments posted online. Their research aimed to assess the feasibility of using automated sentiment analysis tools to analyze patient feedback for healthcare quality improvement. The study found that sentiment analysis could effectively identify positive and negative sentiments in patient comments, providing valuable insights for healthcare providers. The authors also discussed the limitations of current tools in understanding context and the need for further refinement to enhance accuracy.

Furqan Rustam and colleagues (2021) [10] conducted a performance comparison of supervised machine learning models for COVID-19 tweets sentiment analysis. Their study evaluated classifiers such as Random Forest, Support Vector Machine, and Logistic Regression, analyzing their effectiveness in classifying sentiments expressed in COVID-19 related tweets. The research highlighted the challenges of processing social media data, including slang, abbreviations, and the rapid evolution of language. The authors concluded that certain models performed better in capturing the public's sentiment during the pandemic, providing insights for public health communication strategies.

Haiyun Peng, Erik Cambria, and Amir Hussain (2017) [11] reviewed sentiment analysis research in the Chinese language. Their work examined the unique challenges associated with Chinese text processing, such as word segmentation and the handling of idiomatic expressions. The authors discussed various approaches, including lexicon-based methods and machine learning techniques, assessing their applicability to Chinese sentiment analysis. They also highlighted the scarcity of annotated resources and the need for developing language-specific tools to improve analysis accuracy.

3. PROPOSED SYSTEM

The proposed system aims to enhance sentiment analysis by leveraging Recurrent Neural Networks (RNNs) for real-time, multilingual classification in global markets. The methodology unfolds in several key steps: represented as shown below in Figure 1

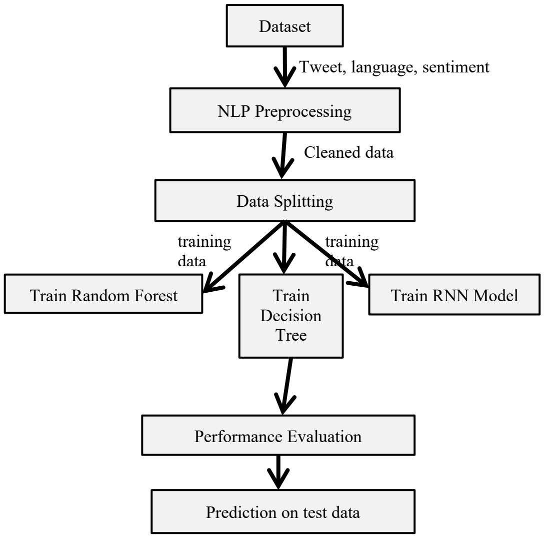

Figure 1: Proposed System

The proposed methodology includes the following key components:
Step 1-Sentiment Dataset: The foundation of the system is a comprehensive sentiment dataset encompassing diverse languages namely English, Spanish, Dutch, etc. The dataset consists of Tweets made in X (formerly known as Twitter), the associated language, and the sentiment expressed in terms of star rating ranging from 0 stars to 5 stars.
Step 2-Text Preprocessing: Prior to analysis, the textual data undergoes rigorous preprocessing. This phase includes tokenization, stemming, and TF-IDF vectorization, which collectively standardize the text, reduce dimensionality, and extract pertinent features, thereby enhancing the quality of input for subsequent modelling.
Step 3-Existing Algorithm: The Random Forest algorithm, an ensemble learning method, constructs multiple decision trees during training and outputs the mode of the classes for classification tasks. It is known for its robustness and accuracy in handling large datasets and mitigating overfitting. The Decision Tree Classifier **is** a supervised learning algorithm that recursively splits the dataset into subsets based on feature conditions, forming a tree-like structure for classification. It is effective for interpretable decision-making and works well with both categorical and numerical data but may overfit without proper pruning.
Step 4-Proposed Algorithm: The proposed system employs Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, to capture temporal dependencies in text data. RNNs are adept at processing sequences, making them ideal for analysing the context and sentiment in textual information.
Step 5-Performance Comparison: A comparative analysis is conducted between the existing Random Forest model and the proposed RNN-based model. This evaluation assesses metrics such as accuracy, precision, recall, and F1-score to determine the efficacy of the RNN in real-time, multilingual sentiment classification.
Step 6-Prediction on Test Data: The model is given test data and it accordingly predicts the sentiment associated with the specific tweet.
After the dataset is loaded, text preprocessing is done. The steps involved are lowercasing, removal of stop words and special characters, lemmatization and stemming. To tokenise the text we have used Keras Tokenizer. Then padding of sequences is done to ensure uniform input size. The target variable (sentiment) is encoded into numerical values (categorical encoding). This ensures the labels are in a format that the neural network can process. The data is split to 80% training data and 20% testing.

Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are a class of artificial neural networks designed to process sequential data by maintaining a form of memory through their internal states. This capability makes them particularly effective for tasks involving time-series data, natural language processing, and other applications where the order and context of inputs are crucial.
RNNs process data sequentially, with each step's output influenced by the previous computations. At each time step, the network receives an input and updates its hidden state, which serves as a memory of past inputs. This hidden state is then used to produce the output for that time step. The recurrent connections allow the network to capture temporal dependencies, enabling it to learn patterns over time. However, traditional RNNs can struggle with long-term dependencies due to issues like vanishing gradients. Variants such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) have been developed to address these challenges.
The architecture of an RNN consists of an input layer, one or more hidden layers with recurrent connections, and an output layer. The hidden layers are interconnected, allowing information to persist across time steps. Each hidden unit receives input from the previous time step's hidden state and the current input, processes this information, and passes the result to the next time step. This structure enables RNNs to maintain context and learn temporal patterns.

Advantages:

- **Sequential Data Processing**: They are inherently suited for tasks involving sequences, such as language modeling, speech recognition, and time-series forecasting.
- **Contextual Understanding**: RNNs can capture the context of inputs over time, making them effective for tasks where the meaning depends on the sequence of inputs.
- **Parameter Sharing**: The use of the same weights across all time steps reduces the number of parameters, making the model more efficient and less prone to overfitting.
- **Flexible Input and Output Lengths**: RNNs can handle inputs and outputs of varying lengths, which is beneficial for tasks like machine translation and speech recognition.
- **Adaptability**: They can learn from data where the temporal dynamics are complex and non-linear, allowing them to model a wide range of sequential patterns.

4. EXPERIMENTAL ANALYSIS

The implementation of the Dynamic Sentiment Classifier Using Recurrent Neural Networks (RNNs) involves a systematic approach to analyze and classify sentiment in real-time across multiple languages for global markets. The system is designed to handle large volumes of unstructured text data, preprocess it efficiently, and apply advanced machine learning and deep learning models to achieve accurate sentiment classification. Below is a detailed description of the implementation process:
Structure of the Dataset: The dataset is structured in a tabular format with three primary columns:

- **Tweet:** This column contains the text data, which includes tweets, reviews, or short comments. The text is written in multiple languages, reflecting the diversity of global communication.

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

- **Language:** This column specifies the language of the corresponding tweet. The dataset includes text in languages such as English (en), Spanish (es), and French (fr), among others.

- **Sentiment:** This column represents the sentiment label associated with each tweet. The sentiment is rated on a scale of 1 to 5 stars, where: **1 star:** Extremely negative sentiment, **2 stars:** Negative sentiment, **3 stars:** Neutral sentiment, **4 stars:** Positive sentiment, **5 stars:** Extremely positive sentiment.

Multilingual Nature: The dataset is multilingual, containing text in various languages. This feature is critical for training models that can analyze sentiment across different linguistic and cultural contexts. For example:

- **Spanish (es):** Tweets in Spanish discuss topics such as sports, public infrastructure, and personal experiences.

- **English (en):** English tweets include book reviews, personal opinions, and general commentary.

- **French (fr):** French text covers topics like internet popularity and market trends.

Sentiment Labels: The sentiment labels in the dataset are assigned on a 5-star scale, providing a granular representation of sentiment intensity. This scale allows the system to distinguish between subtle differences in sentiment, such as:

- **Neutral Sentiment (3 stars):** Text that expresses neither strong positive nor negative emotions.

- **Positive Sentiment (4 stars):** Text that conveys satisfaction or approval.

- **Extremely Positive Sentiment (5 stars):** Text that expresses strong enthusiasm or admiration

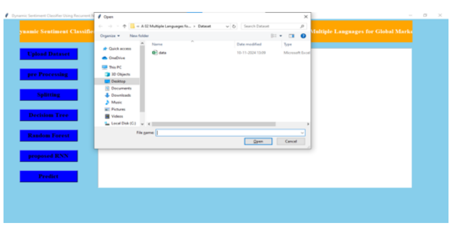

Figure 2: Upload Dataset

As the above figure 2 shows, first the dataset is uploaded. This dataset consists of tweets in various languages.

Figure 3: Preprocessing and Splitting Data

As the above figure 3 shows, next we preprocess the data. The text data is preprocessed by removing stop words, text lemmatization, stemming, etc.

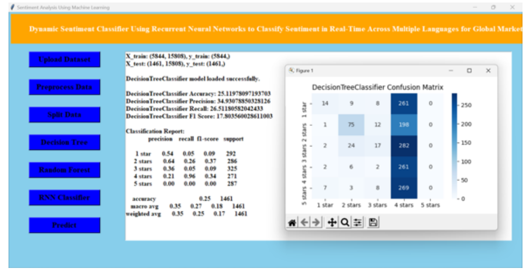

Figure 4: Decision Tree Classifier

As the above figure 4 shows, upon clicking Decision Tree Button, the metrics are displayed and the confusion matrix is shown.

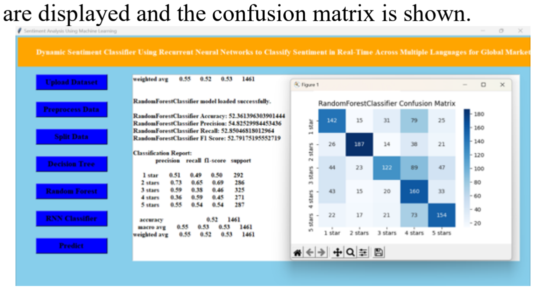

Figure 5: Random Forest Classifier

As the above figure 5 shows, upon clicking Random Forest Button, the metrics are displayed and the confusion matrix is shown.

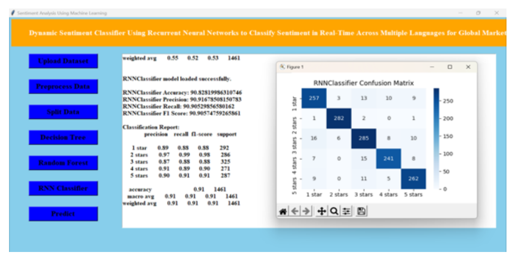

Figure 6: RNN Classifier

As the above figure 6 shows, upon clicking RNN Classifier Button, the metrics are displayed and the confusion matrix is shown.

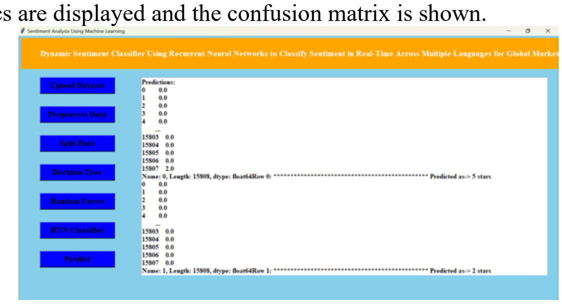

Figure 7: Predict

The above figure 7 shows, a **Sentiment Analysis Application** that processes datasets using machine learning models. The output displays predicted sentiment scores, indicating sentiment variations. Finally, these scores are converted into star ratings, such as **5 stars** and **2 stars**.

5. CONCLUSION

The project titled Dynamic Sentiment Classifier using Recurrent Neural Networks to Classify Sentiment in Real-Time Across Multiple Languages for Global Markets has effectively addressed the complexities involved in multilingual sentiment analysis. Leveraging Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU), the developed system demonstrated significant proficiency in Capturing temporal dependencies and contextual nuances in diverse text data.

The end-to-end project pipeline encompassed crucial stages including data collection, extensive data preprocessing, handling class imbalances with the Synthetic Minority Over-sampling Technique (SMOTE), model training, and comprehensive evaluation. Comparative studies revealed that the proposed LSTM and GRU models outperformed traditional machine learning algorithms such as Random Forest and Decision Tree classifiers in terms of key performance metrics: accuracy, precision, recall, and F1-score.

Key accomplishments of the project are achieved high accuracy in real-time sentiment classification across multiple languages, enhanced the model's capability to manage code-switching and context-sensitive expressions, developed a scalable architecture capable of processing large volumes of unstructured data efficiently, and enabled actionable business insights to improve decision-making, customer engagement, and marketing strategies.

This project represents a significant advancement in the domain of sentiment analysis, highlighting the potential of deep learning models for real-time, multilingual data processing. By combining robust architectures with comprehensive preprocessing techniques, the system offers practical solutions to real-world sentiment classification challenges. Continued enhancements focusing on scalability, multimodal integration, and ethical AI practices will broaden the system's applicability in diverse fields such as market research, public sentiment monitoring, and customer service improvement.

REFERENCES

[1]. Indhraom, Prabha, M., G., Umarani, Srikanth. (2019). Survey of Sentiment Analysis Using Deep Learning Techniques. 1-9. doi: 10.1109/ICIICT1.2019.8741438

[2]. Indhraom, Prabha, M., G., Umarani, Srikanth. (2019). Survey of Sentiment Analysis Using Deep Learning Techniques. 1-9. doi: 10.1109/ICIICT1.2019.8741438

[3]. M., Nivaashini., R., S., Soundariya., P., Thangaraj. (2018). Comparative Analysis of Machine Learning Approaches for Twitter Sentiment Analysis. Journal of Computational and Theoretical Nanoscience, 15(5):1743-1749. doi: 10.1166/JCTN.2018.7371

[4]. Mariam, Biltawi., Wael, Etaiwi., Sara, Tedmori., Amjad, Hudaib., Arafat, Awajan. (2016). Sentiment classification techniques for Arabic language: A survey. 339-346. doi: 10.1109/IACS.2016.7476075

[5]. Muhammad, Rehan., Furqan, Rustam., Saleem, Ullah., Safdar, Hussain., Arif, Mehmood., Gyu, Sang, Choi. (2021). Employees reviews classification and evaluation (ERCE) model using supervised machine learning approaches. Journal of Ambient Intelligence and Humanized Computing, 1-18. doi: 10.1007/S12652-021-03149-1

[6]. Ronglei, Hu., Lu, Rui., Ping, Zeng., Lei, Chen., Xiaohong, Fan. (2018). Text Sentiment Analysis: A Review. doi: 10.1109/COMPCOMM.2018.8780909

[7]. B, Schuller., J-G, Ganascia., Laurence, Devillers. (2016). Multimodal Sentiment Analysis in the Wild: Ethical considerations on Data Collection, Annotation, and Exploitation.

[8]. Basant, Agarwal., Namita, Mittal. (2014). Machine Learning Approaches for Sentiment Analysis. 1740-1756. doi: 10.4018/978-1-4666-6086-1.CH011

[9]. Felix, Greaves., Daniel, Ramirez-Cano., Christopher, Millett., Ara, Darzi., Liam, Donaldson. (2013). Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online. Journal of Medical Internet Research, 15(11) doi: 10.2196/JMIR.2721

[10]. Furqan, Rustam., Madiha, Khalid., Waqar, Aslam., Vaibhav, Rupapara., Arif, Mehmood., Gyu, Sang, Choi. (2021). A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. PLOS ONE, 16(2) doi: 10.1371/JOURNAL.PONE.0245909

[11]. Haiyun, Peng., Erik, Cambria., Amir, Hussain. (2017). A Review of Sentiment Analysis Research in Chinese Language. Cognitive Computation, 9(4):423-435. doi: 10.1007/S12559-017-9470-8