Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.031321
Article

Tech Science Press

A Cloud Based Sentiment Analysis through Logistic Regression in AWS Platform

Mohemmed Sha*

Department of Software Engineering, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, AlKharj, Saudi Arabia
*Corresponding Author: Mohemmed Sha. Email: ms.mohamed@psau.edu.sa
Received: 14 April 2022; Accepted: 17 May 2022

Abstract: The use of Amazon Web Services is growing rapidly as more users are adopting the technology. It has various functionalities that can be used by large corporates and individuals as well. Sentiment analysis is used to build an intelligent system that can study the opinions of the people and help to classify those related emotions. In this research work, sentiment analysis is performed on the AWS Elastic Compute Cloud (EC2) through Twitter data. The data is managed to the EC2 by using elastic load balancing. The collected data is subjected to preprocessing approaches to clean the data, and then machine learning-based logistic regression is employed to categorize the sentiments into positive and negative sentiments. High accuracy of 94.17% is obtained through the proposed machine learning model which is higher than the other models that are developed using the existing algorithms.

Keywords: Sentiment analysis; AWS; cloud; data analytics; machine learning

1 Introduction

Cloud computing plays a major role by offering on-demand applications as per the requirement. These service models are used by the service providers to simplify the process of delivering services directly to the clients [1]. It is seen that cloud infrastructure performance is strongly reliant on productive-backed policies and smart choices by the providers and customers. Also, other characteristics of cloud technology consist of load balancing, autoscaling, latency, Service Level Agreements, energy usage, response time, execution speed, deadline limit, conversion, make span, and optimization are considered by customers and vendors to sustain the Quality of Service (QoS). The approaches such as metaheuristic are implemented to streamline certain QoS and resolve the related combinatorial problems in the cloud-based setup where traditional algorithms find it difficult to find the optimal solutions efficiently and effectively. Comparably, these algorithms incorporate quick decision-making and easy convergence [2]. Hence, to explore the key advantages of cloud computing the advancement of heuristics, metaheuristic, hybrid metaheuristic, and applied machine learning techniques are used. To improve the detection performance and load balancing of a machine learning model metaheuristic algorithms are implemented. Amazon Web Services (AWS) has massive advantages over the contemporary cloud. AWS cloud has benefits to provide in terms of data

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

security, regulatory enforcement, measurability, scalability, cost-effectiveness, various infrastructure, automotive scaling, data connectivity at all times, data-centric privacy, and high-performance computing [3].

Sentimental analysis is the computational study to understand people's sentiments, thoughts, opinions, and feelings on a particular product or service [4]. It has been studied a lot by many researchers. However, there is very little research on sentiment analysis with respect to cloud services. To find out what some feel was a vital aspect of the conduct of knowledge collection. Sentiment analysis in Natural Language Processing (NLP) is a significant area of research focusing on the extraction of meaning from textual information. The area of research is sentiment analysis, which collects and analyses the thoughts, feelings, sentiments, and behaviors of persons, programs, activities, goods, and instructors and their characteristics or attributes towards individuals and organizations [4]. There are also several activities of various titles, such as opinion mining, sentimental research, emotional mining, opinion mining, subjectivity analyses, and mining evaluation. Nonetheless, both are now examined in a sentimental study or opinion mining framework. The first sentimental study and the term opinion mining were used around two decades earlier [5,6].

Sentiment analysis is a common research problem in Natural Language Processing, which is quite difficult to achieve. As more text is accessible in Digital formats to be one of the most important study fields in NLP, the field of sentiment analysis has been growing rapidly since 2000. Duo with the broad variety of technologies and the wide realistic use of different markets, in almost any field of social networking, policy and marketing analyses. Internet networking and technologies (e.g., comments, ratings, messages, news applications, sharing to social media sites) are growing massively, with the increased usage of content by individuals and organizations of those media in decision-making. Simultaneous with social network development is the fast increase in the sentimental study in science. Studying in the field of emotion analysis also has a significant effect on other fields, such as management science, political science, economics, and social sciences, as they are all influenced by the views or feelings of the citizens. Sentiment research may generally be carried out at three levels: text stage, sentence level, component, or aim level [7].

In this research sentiment analysis is achieved on the AWS Elastic Compute Cloud (EC2) through Twitter data. The data is managed to the EC2 by using elastic load balancing. The collected data is subjected to preprocessing approaches to clean the data, and then machine learning-based logistic regression is utilized to classify the sentiments of people into positive and negative sentiments.

2 Literature Review

Cloud service companies and customers encounter many obstacles in speeding up connectivity to cloud technology. The retrieval of only useful details is a challenging task that demands more time, given the vast amount and variety of data put in the cloud [8,9]. The problem becomes more distressing as big, computer-complicated, and resource-intensive programs need to be processed. Data pre-processing in such situations could play an important role in reducing execution time and memory requirements during an offline classification of knowledge for machine learning models during online processing. To maintain optimum load balancing, the job assignment for VMs must also be carried out carefully. Elastic Compute Cloud of AWS, shortly known as EC2 is a platform used in the cloud for users requiring virtualization [2]. The platform can be rented to the required users for running certain programs.

Hybrid metaheuristics are used to categorise and balance loads, fault tolerances, costing analyses, energy savings, and for several other purposes. However, a recent state of the art is the grouping of cloud data into different file formats which is already used in AWS [3,10]. However, the classification of data files in cloud storage, such as audio, video, documents, pictures and charts, involves extra work to accurately classify and balance the load, which can be solved in two stages. In the first stage, a classification algorithm needs to be

built that correctly classifies the cloud data sets, causing erroneous data groups. In the second stage, the resulting data class is inserted into an algorithm such as metaheuristic load balancing.

While various types of applications can be built into the cloud, sentiment analysis is something that is not fully explored in the cloud space. In natural language processing, sentiment analysis is concerned with understanding information in the text. It is also known by another term called opinion mining, which is the discovery of an individual, service's sentiments, and attributes [11,12]. The term sentiment analysis was first developed in Nasukawa et al. [5], however, the concept existed previously. Since the year 2000, sentiment analysis has developed significantly as the digital text has become abundant, especially within NLP studies.

This is due to the variety of applications in industry and a large amount of practical use in virtually every area, like social media, political analyses, and marketing. The huge rise in social networking and technologies with the growing usage of advertising in these media by individuals and organisations. Simultaneously with that of social media, the exponential development in opinion analysis in the study is. Therefore, studies in the domain of sensitivity analyses have a great influence on other fields outside the NLP since they are both influenced by the views or feelings of individuals, such as management science, political science, economics, and social science [13].

In general, an interpretation of sentiment may be carried out at three levels: text stage, sentence level and aspect or destination level [14]. The study of text and sentence levels helps to attribute the polarity of the emotions to a defined opinion at different granular levels. The goal is to detect sentimental polarity towards the opinions of targets provided in a sentence, using Target Based Sentiment Analysis (TBSA). Neither the level of a text nor the level of a sentence figures out just what people liked and liked. TBSA has a more detailed analysis, and its results are more accurate [8].

TBSA remains daunting, though, as it is difficult to model the conceptual connection between an aim and its surrounding meaning terms. Furthermore, common meaning terms affect the polarity of a sentence against the goal words. The semantic relation between the goal and contextual words must therefore be captured when constructing a model of TBSA learning. TBSA or Aspect-based feeling analysis was early referred to as feature-based opinion mining and summary. The aspect level looks at the opinion itself explicitly, instead of taking the feeling over the whole article, paragraph, or word. It is focused on the premise that an opinion caries a feeling (positive or negative) and an objective [15].

Rezaeinia et al. [16] have performed opinion mining on pre-trained data. Vectorization requires a substantial amount of data for building the necessary vectors before training. The proposed method has made use of an enhanced wave vector based on a lexicon, where a portion of the speech is obtained, and the words are interpreted through the proposed algorithm. Song et al. [17] have performed opinion mining for both images and videos. The spatial regions of the images have been positioned and augmented before training the Convolution Neural Network (CNN) classifier. This approach is shown to be more effective when the data is placed in multiple layers. The accuracy is even better in case textual data has accompanied the images. Hence, the researchers have a semi-supervised approach since it can work with both labeled and unlabeled data.

Support Vector Machine (SVM) is a versatile algorithm that can perform classification and regression tasks, which allows it simpler to identify. This means the components of the data set each with "n" dimensions divided by a maximum margin called a hyperplane. Ant Colony Optimization (ACO) provides better performance in load balancing challenges, and with a variety of variations, it is one of the most commonly employed algorithms. ACO is resilient and will find solutions more quickly [18] Moreover, this can be rapidly combined with other metaheuristics as ACO look at the same time. This helps achieve successful results. As a result of its diversity, ACO has been used in several different studies including controlled models for instruction, such as the rules on categorization [7,19].

Zhang et al. [20] have executed opinion extraction for sentiment analysis. Different reasoning for the sentiments has been used to identify the sentiments. NLP has been utilised to gather the previous opinions. Long Short Term Memory (LSTM) has been employed for the bidirectional and multilayered encoding process. Transitional findings are critical for decoding; global normalisation approaches and beam search have been used here. Logistic Regression has been implemented by Majumder et al. [21], for identifying the sentiments of the people during the COVID pandemic. The researcher has also implemented the same using SVM. Data has been collected from Twitter for the period between March and June 2020, and then the polarity of the tweets has been classified into positive, negative, and neutral. Python programming has been used as a platform to classify these tweets, and high accuracy has been obtained for the logistic regression model.

While there are many models for the analysis the logistic regression model seems to have high accuracy and can be implemented in this research. While the existing studies have implemented the work in Python, this work implements the code in the AWS cloud for direct and instant sentiment analysis.

3 Logistical Regression

Logistic regression is a suitable regression analysis for the care of the dependent variable (binary). It is a statistical analysis, like other regression analyses. It is used to define and illustrate the relationships of one binary variable based on one or more individual variables, whether nominal, ordinal, interval, or ratio-level [22]. Two forms of classification models are commonly available: generative and discrimination models. Broadly, the probability of features is calculated by both models. The major distinction is that the generative model initially models the distribution of mutual probability and then calculates the dependent probability, whereas a discriminatory model explicitly models the distribution of probability.

Another significant factor is the model fit when choosing the model for the logistic regression analysis. In addition, the sum of variance explicit in log-odds will often increase when independent variables are applied to a logistic regression model [23]. The addition of more and more variables to the model, however, will lead to an overflow that decreases the generalization of the model over the data the model is fit to provide.

4 Sentiment Analysis

A relevant dataset is necessary to perform the implementation. The data that is relevant to this analysis is collected from Twitter Tweepy. Twitter always remains the exact indicator of the pulse of the people to indicate what is happening with different matters. Hence, this dataset is considered here for the emotional analysis of textual data. The tweets are not collected based on a specific topic, but rather collected randomly, hence they may contain any type of data. A total of 49,159 rows of data is obtained from the tweets. From this data, it is evaluated whether the tweets are of positive or negative sentiments. This dataset must contain reviews or comments regarding a certain issue which indicates that the customer or the user will have some form of feelings or emotions in the feedback. This collected data will include plenty of irrelevant data that must be removed, therefore relevant pre-processing algorithms and appropriate classifiers must be studied and chosen. The emotional analysis will be done dependent on the kind of optimisation approach. Therefore, the ideal combination of optimisation approaches must be determined. After that, the accuracy is compared with that of existing approaches. In this study, logistical regression is used for classifying the sentiments. To train and test the model the dataset is divided into a training and testing dataset at a ratio of 65:35 with 31962 tweets for the training dataset, and 17197 tweets for the test dataset. Some rows of the training dataset are shown in Fig. 1.

	id	label	tweet
0	1	0	@user when a father is dysfunctional and is s...
1	2	0	@user @user thanks for #lyft credit i can't us...
2	3	0	bihday your majesty
3	4	0	#model i love u take with u all the time in ...
4	5	0	factsguide: society now #motivation

Figure 1: Training dataset

The collected data must be processed before the classification procedure. As a result, pre-processing is required. Primarily, it needs to be guaranteed that the information in the dataset that make sense linguistically. Hence, lexical semantic and rule-based approaches are utilised at this step for retrieving the features from the original text. To achieve this the SentiWordNet database is utilised at this step for all the linguistic parsing in the content. Reviews are cleaned of any extraneous information, such as Uniform Resource Locators (URLs) and slang terms, to improve their categorization. The length of the tweets may be useful for identifying the sentiments, since longer tweets may have more aspects or targets. Hence, the length of each tweet is identified and added as a separate column as shown in Fig. 2. The sentence's length may be an essential consideration in the training process. The text provides all the relevant information. If there is a big phrase, just the significant bits must be retrieved and not the full paragraph. Shorter reviews, on the other hand, will include fewer features, making feature extraction unnecessary.

	id	label	tweet	len
0	1	0	@user when a father is dysfunctional and is s...	102
1	2	0	@user @user thanks for #lyft credit i can't us...	122
2	3	0	bihday your majesty	21
3	4	0	#model i love u take with u all the time in ...	86
4	5	0	factsguide: society now #motivation	39
5	6	0	[2/2] huge fan fare and big talking before the...	116
6	7	0	@user camping tomorrow @user @user @user @use...	74
7	8	0	the next school year is the year for exams.ð□□...	143
8	9	0	we won!!! love the land!!! #allin #cavs #champ...	87
9	10	0	@user @user welcome here ! i'm it's so #gr...	50

Figure 2: Length of tweets column

The training data contains a label on whether the tweet has a positive or negative sentiment. Here, 1 represents positive sentiment, and 0 represents negative sentiment. Most of the tweets in the training dataset contain negative sentiments with 29720 tweets with negative sentiments, and 2242 tweets with positive sentiments as shown in Fig. 3.

There are however some limitations in applying the machine learning approaches as they cannot conceive organically like people, but these algorithms can compute based on some training on the split dataset. As a result, further data modifications are applied to better train the system. Based on the matrix of features, the sentences are translated into vectors. Some of these adjustments may are aimed at developing new vectors which retain the data in a more efficient fashion. This is accomplished by making certain that the dataset has no errors. This vectorization reflects the tokenization of the tweets, which is accomplished using the genism model.

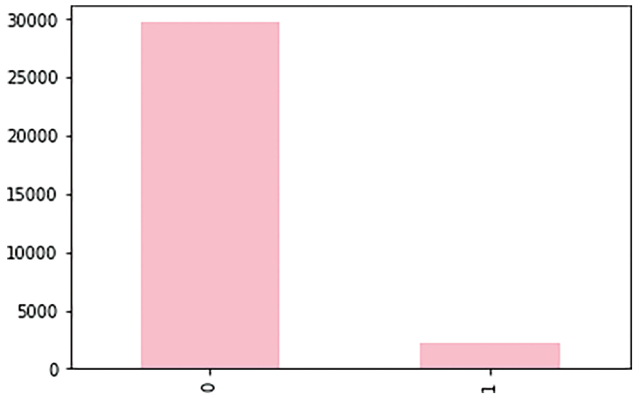

Figure 3: Positive and neutral comments in training dataset

Another characteristic is that it is case-sensitive. It is required to convert the content to a single typeface since the reviews may include both upper- and lower-case letters. This is because the cases may be replaced and may be interpreted differently by the incremental Artificial Intelligence (AI) algorithm. Therefore, they may be preserved in a different way when they are turned into some numerical value. Therefore, in this study, the complete text data is changed into lower case regardless of the originating data. Since their existence is not very significant, the punctuation is left just as it is. The tweets are then analysed to identify the most used words in the tweet. This is done to identify the intercorrelation between each of them. The top 30 commonly used words are shown in Fig. 4. The word "user" appears in 17,500 tweets and is the most used word. This is because the dataset has converted all the username mentions into "user", hence anyone who has mentioned someone else in their tweets has been replaced with "@user". Hence, the user is not a word that is used directly by the people. When this word is ignored, the other common words are love, day, amp, and happy where they appear in roughly 2,500 words each.

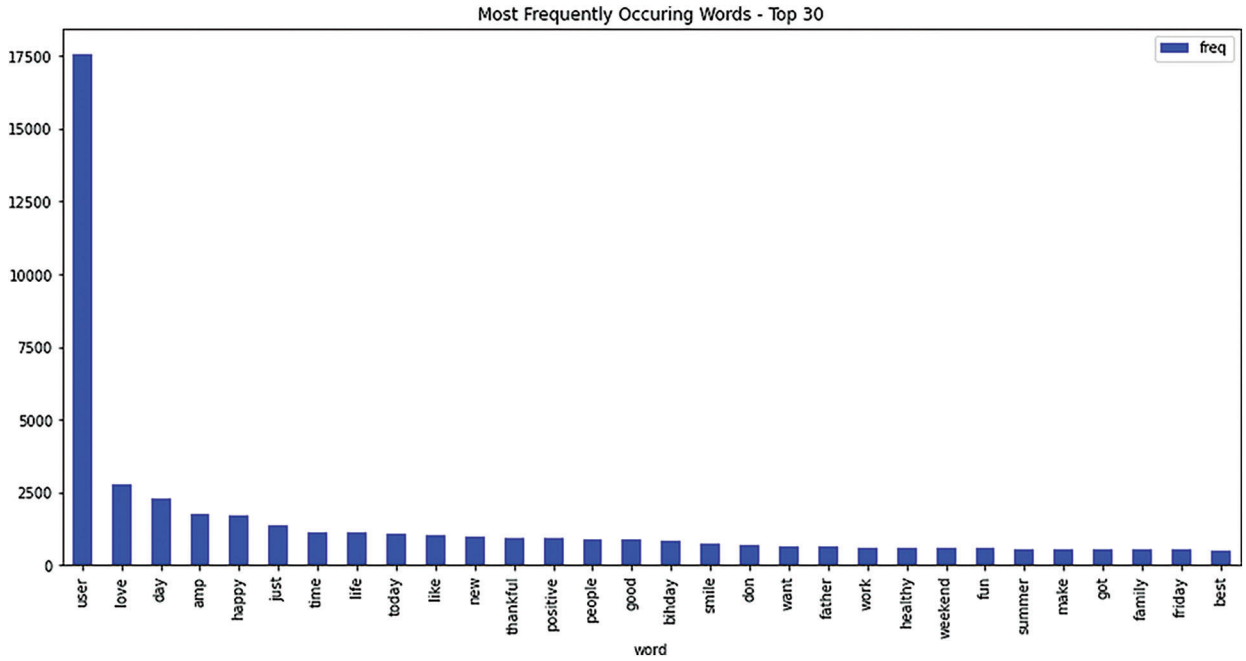

Figure 4: Top 30 most frequently occurring words

This frequency of the commonly used words is represented graphically in a word cloud format, where the frequent words are shown larger. The word cloud representation is shown in Fig. 5.

Figure 5: Word cloud representation

This frequency of the commonly used words is represented graphically in a word cloud format, where the frequent words are shown larger. The word cloud representation is shown in Fig. 5 and may contain a mixture of both positive and negative sentiment words. The positive and negative words are separated in a word cloud format and shown in Figs. 6 and 7. It is worth noting that positive and neutral are clubbed together. Also, some words may have both aspects and are repeated in both the classification since their adjectives are responsible for being positive or negative.

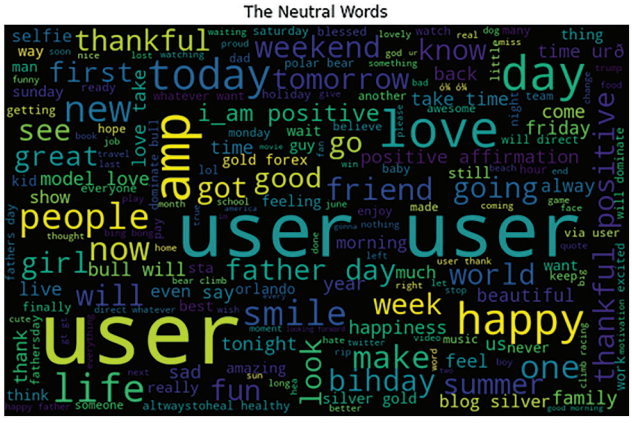

Figure 6: Word cloud representation for neutral words

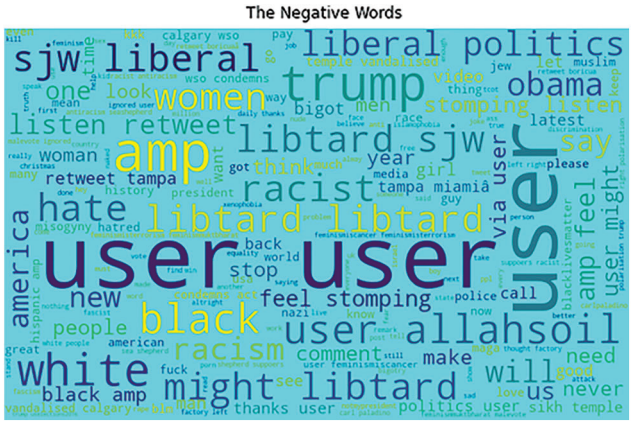

Figure 7: Word cloud representation for negative words

The pre-processed phrases are analysed to determine the type of sentiment in the final step. The analysis classifies data using the proposed algorithm that is incremental in nature. The nouns are highlighted and extracted from the sentences. These words' feelings need to be discovered and categorised. This study uses Logistic Regression to extract the feelings. The hybrid classifier is trained using the labeled training data. This can be done when each unlabeled piece of data can be run, and its characteristics are gathered. The thoughts of individual phrases may also be gathered by running the sentences one at a time. The outputs include the sentiments and their polarity.

5 Integration with AWS

The sentiment analysis performed in the previous section is connected to a cloud Application Programming Interface (API). The API is first created, and the trained model is fed into the server. When the user wants to find the sentiments of a new tweet, the tweet is fed into the API, and gets back the sentiment of the tweet. These sentiments can be obtained either in form of a plain text file or in form of a Comma Separated Values (CSV) file through the data ingestion script. Data ingestion is a mechanism by which data is transferred to a destination from one or more origins and may be saved and evaluated further. Data can be available in multiple formats from multiple centers, including Relational Database Management Systems (RDBMS) and CSVs [24, 25]. Here CSV files are used in this research. The ingestion script is performed using the Amazon Kinesis Data Firehose platform which makes it easier for the data to stream freely [26].

The implementation is performed on the EC2 environment. Nevertheless, this does not limit the results, as there are lots of reports of performance metrics for Single-Job benchmarks. For this implementation, consistent environments with many cores are selected. Amazon EC2 presents a variety of options depending on the size and processing speed requirements. In this research work, the images listed in support of the 32 and 64-bit instances are used, which are mainly based on a Fedora Core such as along with a Linux kernel, that is compatible with Python programming. The virtual images utilized for the benchmarks in addition have a functioning pre-configured Message Passing Interface (MPI). Moreover, the benchmarks are compiled utilizing GNU C/C++ with oops command-line arguments. No additional architecture or depending on optimizations are required for this sentiment analysis.

Based on the obtained sentiments, the user can make necessary decisions. The user may use the sentiment analysis for various applications including the opinion of customers, the opinion of the public, or the opinion of the general spectators. The final output from the model is obtained in form of the (JavaScript Object Notion (JSON) format. The complete flow of the implementation is shown in Fig. 8.

Since the model uses AWS EC2, elastic load balancing is also used in this research. Since the data is captured from Twitter in real-time, the traffic may be huge and must be controlled. The Elastic Load Balancing will be able to manage these incoming tweets and decide to meet the traffic demands. Only a limited amount of load must be allowed to enter the AWS model due to space restrictions, hence the elastic load balancing is used in this model and will act as a single point of contact to the user [27]. It will scale the data such that the available cloud space is utilized to the maximum. If an EC2 instance is unsuccessful, then the load balancing automatically reallocates the incoming data to an available EC2 instance.

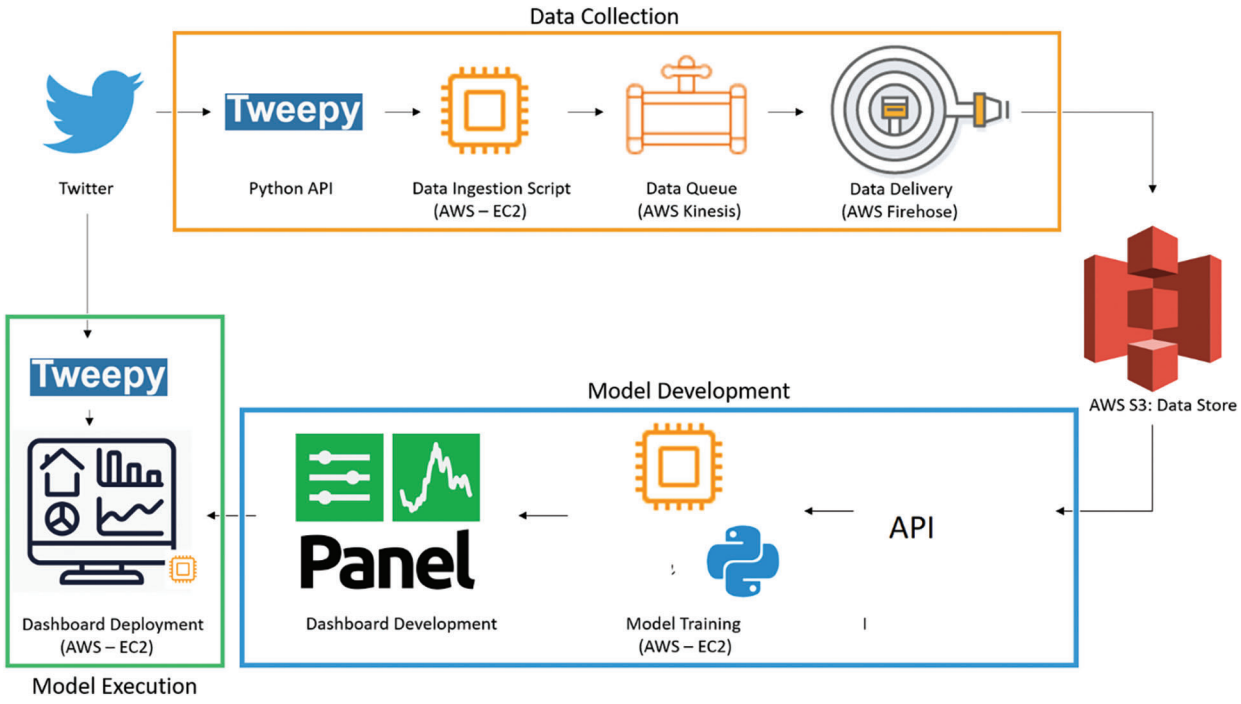

Figure 8: Flow of implementation

Performance Metrics

The sentiment analysis implementation is run, and the parameters like true positive, true negative, false positive, and false negatives are obtained as shown below. The true negative obtained is 7185, while the true positive obtained is 340.

7185 247
219 340

While the training accuracy of 98.51% is obtained, a high testing accuracy of 94.17% has been obtained as shown in Fig. 9.

```
Training Accuracy : 0.9851487213716574
Validation Accuracy : 0.9416843949443123
f1 score : 0.5933682373472949
```

Figure 9: Performance metrics

To provide a clearer insight various performance metrics such as recall, f1 score, precision, and accuracy are calculated from the confusion matrix and given below

$$\text{Precision} = \frac{TP}{TP + FP} = 0.5792$$

$$\text{Recall} = \frac{TP}{TP + FN} = 0.6082$$

$$\text{F1 Score} = 2 * \frac{Precision * Recall}{Precision + Recall} = 0.5933$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = 0.9417$$

Tab. 1 compares the accuracy of the proposed method with that of other existing approaches. It is seen that the proposed approach has 94.17% accuracy, which is very much higher than the other approaches. Hence, the proposed algorithm has superior performance. In addition, since this implementation utilizes direct twitter data with the help of AWS, the work has a large significance in identifying the sentiments immediately in real-time.

Table 1: Comparison of accuracy of different algorithms

Technique	Accuracy
Neural Network (NN) [26]	71%
Convolutional Neural Network (CNN) [28]	74.5%
Ensemble – Naïve Bayes, Random Forest, SVM, Logistic Regression [29]	85.83%
Proposed LSTM-RNN technique	94.17%

6 Conclusion

Sentimental analysis has been implemented in this research using machine learning through logistic regression. While many researchers have used ordinary CSV data for analysing the sentiments, the possibility of analysing live tweets was explored by integrating the model into AWS. EC2 was used for successful integration, and elastic load balancing is used to regulate the live Twitter data. Pre-processing text categorization is the most prevalent problem encountered by researchers in this work. However, this was performed successfully in this research. These factors are crucial because they affect how many words are taught to the machine learning algorithm. This involves a variety of transformations and pre-processing procedures. Many various elements of the tweets are gathered and sent into the logistic regression classifier as training data. Using the classifier, new tweets are sorted into positive, negative, and neutral sentiment categories. The accuracy of 94.17% is greater than the conventional algorithms, based on the results gathered thus far. By optimising the suggested system in the future, more accurate results are expected in the future.

Acknowledgement: The author would like to thank the Deanship of Scientific Research, Prince Sattam Bin Abdulaziz University, KSA for supporting this project.

Funding Statement: This research project was supported by the Deanship of Scientific Research, Prince Sattam Bin Abdulaziz University, KSA, Project Grant No. 2021/01/17783, Sha M, www.psau.edu.sa.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Y. Duan, G. Fu, N. Zhou, X. Sun, N. C. Narendra *et al.,* "Everything as a service (XaaS) on the cloud: Origins, current and future trends," in *Proc. IEEE 8th Int. Conf. on Cloud Computing*, New York, USA, pp. 621–628, 2015.

[2] M. Kalra and S. Singh, "A review of metaheuristic scheduling techniques in cloud computing," *Egypt. Informatics Journal*, vol. 16, no. 3, pp. 275–295, 2015.

[3] H. Singh, "Cloud Computing and AWS," in *Practical Machine Learning with AWS*, Berkeley, CA: Apress, pp. 3–28, 2021.

[4] B. Liu, "Sentiment analysis: Mining opinions, sentiments, and emotions," in *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Pulishers, pp. 1–367, 2015.

[5] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proc. of 2nd Int. Conf. on Knowledge Capture*, Florida, USA, 2003.

[6] K. Dave, S. Lawrence and D. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in *Proc. 12th Int. Conf. on World Wide Web*, Budapest, Hungary, pp. 519–528, 2003.

[7] L. Zhang, S. Wang and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, pp. 1–39, 2018.

[8] M. Junaid, A. Sohail, A. Ahmed, A. Baz, I. A. Khan *et al.,* "A hybrid model for load balancing in cloud using file type formatting," *IEEE Access*, vol. 8, pp. 118135–118155, 2020.

[9] Amazon Web Services, *What is Amazon EC2*? Washington, USA, 2021. [Online]. Available: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts.html.

[10] Amazon Web Services, *Using a PostgreSQL database as an AWS DMS source*, Washington, USA, 2021. [Online]. Available: https://docs.aws.amazon.com/dms/latest/userguide/CHAP_Source.PostgreSQL.html.

[11] A. Nazir, Y. Rao, L. Wu and L. Sun, "Issues and challenges of aspect-based sentiment analysis: A comprehensive survey," *IEEE Transactions on Affective Computing*, vol. 8, pp. 1–20, 2020.

[12] M. J. I. Razin, M. A. Karim, M. F. Mridha, S. M. R. Rifat and T. Alam, "A long short-term memory (LSTM) model for business sentiment analysis based on recurrent neural network," in *Proc. Sustainable Communication Networks and Application*, Singapore, Springer, pp. 1–15, 2021.

[13] X. L. Luan, F. X. Gong, Z. Q. Wei, B. Yin and Y. T. Sun, "Using ant colony optimization and cuckoo search in AUV 3D path planning," *Software Engineering and Information Technology*, vol. 13, pp. 208–212, 2016.

[14] A. Gupta and R. Garg, "Load balancing based task scheduling with ACO in cloud computing," in *Proc. Int. Conf. on Computer and Applications*, Doha, UAE, pp. 174–179, 2017.

[15] A. Ankit and N. Saleena, "An ensemble classification system for twitter sentiment analysis," *Procedia Computer Science*, vol. 132, pp. 937–946, 2018.

[16] S. M. Rezaeinia, R. Rahmani, A. Ghodsi and H. Veisi, "Sentiment analysis based on improved pre-trained word embeddings," *Expert Systems with Applications*, vol. 117, no. 19, pp. 139–147, 2019.

[17] K. Song, T. Yao, Q. Ling and T. Mei, "Boosting image sentiment analysis with visual attention," *Neurocomputing*, vol. 312, no. 12, pp. 218–228, 2018.

[18] S. Aabro, S. Shaikh, R. A. Abro, S. F. Soomro and H. M. Malik, "Aspect based sentimental analysis of hotel reviews: A comparative study," *Sukkur IBA Journal of Computing and Mathematical Sciences*, vol. 4, no. 1, pp. 11–20, 2020.

[19] J. Ning, C. Zhang, P. Sun and Y. Feng, "Comparative study of ant colony algorithms for multi-objective optimization," *Information*, vol. 10, no. 1, pp. 1–11, 2018.

[20] M. Zhang, Q. Wang and G. Fu, "End-to-end neural opinion extraction with a transition-based model," *Information Systems*, vol. 80, no. 3, pp. 56–63, 2019.

[21] S. Majumder, A. Aich and S. Das, "Sentiment analysis of people during lockdown period of COVID-19 Using SVM and logistic regression analysis," *SSRN Electronic Journal*, vol. 7, no. 5, pp. 1–10, 2021.

[22] K. Shah, H. Patel, D. Sanghvi and M. Shah, "A comparative analysis of logistic regression, random forest and KNN models for the text classification," *Augmented Human Research*, vol. 5, no. 1, pp. 1–12, 2020.

[23] W. P. Ramadhan, S. T. M. T. Astri Novianty and S. T. M. T. Casi Setianingsih, "Sentiment analysis using multinomial logistic regression," in *Proc. Int. Conf. on Control, Electronics, Renewable Energy and Communications (ICCEREC)*, Yogyakarta, Indonesia, pp. 46–49, 2017.

[24] L. M. Qaisi and I. Aljarah, "A Twitter sentiment analysis for cloud providers: A aase study of Azure *vs*. *AWS," in Proc. 7th Int. Conf. on Computer Science and Information Technology*, Amman, Jordan, 2016.

[25] G. Sharma, V. Tripathi and A. Srivastava, "Recent trends in big data ingestion tools: A study," *Research in Intelligent and Computing in Engineering*, Singapore: Springer, pp. 873–881, 2021.

[26] I. Karamitsos, S. Albarhami and C. Apostolopoulos, "Tweet sentiment analysis (TSA) for cloud providers using classification algorithms and latent semantic analysis," *Journal of Data Analysis and Information Processing*, vol. 7, no. 4, pp. 276–294, 2019.

[27] Amazon Web Services, *Amazon Kinesis Data Firehose*, Washington, USA, 2021. [Online]. Available: https://aws.amazon.com/kinesis/data-firehose/.

[28] S. Liao, J. Wang, R. Yu, K. Sato and Z. Cheng, "CNN for situations understanding based on sentiment analysis of twitter data," *Procedia Computer Science*, vol. 111, pp. 376–381, 2017.

[29] M. Rezwanul, A. Ali and A. Rahman, "Sentiment analysis on twitter data using KNN and SVM," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, pp. 19–25, 2017.