

Tweet Sentiment Analysis (TSA) for Cloud Providers Using Classification Algorithms and Latent Semantic Analysis

Ioannis Karamitsos¹, Saeed Albarhami¹, Charalampos Apostolopoulos²

¹Rochester Institute of Technology, Dubai, UAE

²Strathclyde Business School, Glasgow, UK

Email: ixkcad1@rit.edu, sat5006@rit.edu, charalampos.apostolopoulos@strath.ac.uk

How to cite this paper: Karamitsos, I., Albarhami, S. and Apostolopoulos, C. (2019) Tweet Sentiment Analysis (TSA) for Cloud Providers Using Classification Algorithms and Latent Semantic Analysis. *Journal of Data Analysis and Information Processing*, 7, 276-294.

<https://doi.org/10.4236/jdaip.2019.74016>

Received: August 30, 2019

Accepted: November 3, 2019

Published: November 6, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The availability and advancements of cloud computing service models such as IaaS, SaaS, and PaaS; introducing on-demand self-service, auto scaling, easy maintenance, and pay as you go, has dramatically transformed the way organizations design and operate their datacenters. However, some organizations still have many concerns like: security, governance, lack of expertise, and migration. The purpose of this paper is to discuss the cloud computing customers' opinions, feedbacks, attitudes, and emotions towards cloud computing services using sentiment analysis. The associated aim, is to help people and organizations to understand the benefits and challenges of cloud services from the general public's perspective view as well as opinions about existing cloud providers, focusing on three main cloud providers: Azure, Amazon Web Services (AWS) and Google Cloud. The methodology used in this paper is based on sentiment analysis applied to the tweets that were extracted from social media platform (Twitter) via its search API. We have extracted a sample of 11,000 tweets and each cloud provider has almost similar proportion of the tweets based on relevant hashtags and keywords. Analysis starts by combining the tweets in order to find the overall polarity about cloud computing, then breaking the tweets to find the specific polarity for each cloud provider. Bing and NRC Lexicons are employed to measure the polarity and emotion of the terms in the tweets. The overall polarity classification of the tweets across all cloud providers shows 68.5% positive and 31.5% negative percentages. More specifically, Azure shows 63.8% positive and 36.2% negative tweets, Google Cloud shows 72.6% positive and 27.4% negative tweets and AWS shows 69.1% positive and 30.9% negative tweets.

Keywords

Azure, AWS, Google Cloud, Machine Learning, Sentiment Analysis, Tweets

1. Introduction

Cloud computing services are accelerated by the unprecedented amounts of data being generated by not only people but also machines and things. Cisco GCI [1] estimates that nearly 850 ZB will be generated by people, machines, and things by 2021, up from 220 ZB generated in 2016. It is clear that organizations are considering cloud-computing services for financial reasons, more flexibility and robust scalability to meet their dynamic business needs. The traditional way of buying servers, storage, network equipment and setting up infrastructure at your own datacenter may not be cost efficient and extremely difficult to afford for most of the start-up companies as well as the process of maintaining. Therefore, fast scaling resources available on-premise datacenters are quite challenging task. Cloud computing strongly helps under such circumstances; however, the process of choosing cloud provider, cloud model or the service model needs caution and careful evaluation from different angles to avoid vendor's lock-in, and make the transition process to competitor's service easier; if decided to switch in future.

It is common, before subscribing to any service or purchasing a product, to tend to make an online research about products or services, looking at opinions, reviews, advices, experiences and recommendations from other people, especially previous customers or users, mostly using social media such as: Twitter, Facebook, LinkedIn, etc. According to a survey [2] of more than 2000 American adults, 81% of internet users have used the internet to conduct a research about a product they are thinking about buying, with 20% doing this on the typical day.

Twitter [3] is a popular social media platform on which users post and interact with tweets (messages). Twitter has numerous amounts of tweets, which mainly express opinions about a diversity of topics. These tweets may express valuable feedbacks and attitudes from customers about a specific brand or service.

This paper discusses the methodology of opinion mining on social media platform (Twitter) to discover people's feedback about cloud computing services in general and attempts to find out the attitudes towards each cloud provider such as Azure [4], AWS [5] and Google Cloud [6]. It attempts to address the following questions like: Which is the customers' opinion about existing cloud providers? What is the impact of social media on brands and services reputation? and Can the social media reviews and opinions change the customer preference toward a brand or service?

This paper is structured as follows. In Section 2, related work and associated research on sentiment analysis is presented. Section 3 focuses on the cloud providers used to collect data. Section 4, discusses the authors' proposed seven steps tweet sentiment analysis methodology. Subsections include, discussions about the collection of the tweet data, the cleaning process, the most frequent words, the related polarity and emotion detection and the sentiment analysis. The emotion and vocabulary detection are based on the two classifiers shown; the Decision

Tree-CART and the Random forest ones. In Section 5, the experimental results are being analysed and in the final section, Section 6, the conclusions drawn out of this paper as well as the future work are shown.

2. Related Work

Sentiment Analysis has attracted increasing attention from many research communities such as data mining, machine learning and deep learning. Automated sentiment analysis focuses on analyzing the content of online posts, determining whether they are positive, negative, or neutral, and aggregating the sentiments detected into a single generic score [7]. The Net Sentiment Score computes a ratio of positive and negative mentions of a topic. A research study [8] for 120 bank customers on the impact of customer satisfaction, loyalty and intentions to switch, revealed that customer satisfaction was positively correlated with customer loyalty and negatively correlated with customer intentions to switch. Many organizations tend to find out what do customers think or tweet about their products in order to effectively plan marketing campaigns and try to gain the positive impact of Word-of-Mouth [9]. An important criterion for a good classification procedure is that, not only it produces accurate classifiers (within the data limits) but also provides insight and understanding into the predictive structure of the data [10].

Text mining is another method of automatically extraction of useful and valuable information from different sources like: articles, social media tools and Tweets. The use of an appropriate dictionary, especially a polarity dictionary containing positive and negative expressions is important for this study. A polarity dictionary is a dictionary that assigns a polarity value to each word giving a positive value to word having a positive meaning and a minus value for a word having negative meaning. In principle, the polarity dictionary is created manually. Particular studies tried to automatize the creation of the polarity dictionaries [11] [12] [13] [14].

A machine learning approach [15] was used for sentiment analysis. In this approach, the relationship between features of textual data and a polarity is trained and learned from the associated machine learning model. Then, the polarity of a new-text is tested by the application of the learning model. The authors [15] proposed a two-step classification approach. For the first step, the target is classified to be objective (neutral) or subjective. In the second step, the subjective target is classified as positive or negative.

In a lexicon approach [16] for the sentiment analysis, a polarity dictionary is required that consists of pairs of polarity words with the polarity value as positive or negative. The definition of polarity value in a manual way is very difficult, especially for an enormous number of words. For this reason, the construction of polarity dictionary is mandatory for the sentiment analysis.

In another approach [17], twitter posts are being classified as a distinct service of two functions. Actually, a tweet can be regarded as creation of information (new) or pure distribution (other users' writings or retweet). In regards to emo-

tional divergence and probability or information distribution; the tweets with high emotional divergence have consequently higher probability of retweet.

To this frame, [18] analyzed communication patterns across on-line chatrooms for associated patterns in terms of real time information exchange and emotion. The models of on-line interaction aimed to analyze and compare the emotional triggers in conversations both on-line and in the real world.

Actual results indicated that people did not change their expression patterns between earlier studied of communication approaches. More specifically, the chatrooms do not show the same heated emotional patterns which can be observed in other on-line forms of communication. Effectively, instant reply chats show a more balanced emotion.

Other authors [19] classified tweets based on emoticons and speech parts correlation. More specifically, emoticons [20] were used for the determination of the tweet sentiment, since, the use of characters limit makes it more appropriate to be on a single sentiment. Tweet remainder, by the use of the Tree Tagger algorithm was split in distinct speech parts [21], indicating which speech part has the highest impact on the overall sentiment.

For large amount of Twitter data, [22] discussed the Perception algorithm for the sentiment classification in the data stream with static selection features.

3. Cloud Providers Background

The National Institute of Standards and Technology (NIST) [23] defines Cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

There are three cloud service models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). Each cloud service model provides a level of abstraction that reduces the efforts required by the service consumer to build and deploy systems. NIST (**Figure 1**) [24] [25] defines the three service models of cloud services as follows:

- Software as a Service (SaaS). The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface.
- Platform as a Service (PaaS). The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider.
- Infrastructure as a Service (IaaS). The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications.

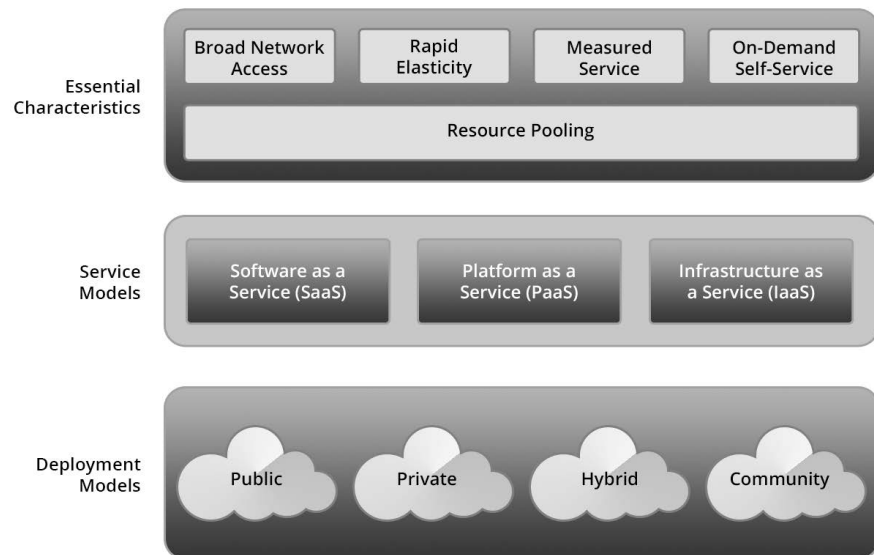


Figure 1. NIST definition of cloud computing (Source: [24]).

3.1. Amazon Web Services (AWS)

Amazon Web Services (AWS) [5] is a cloud services platform from Amazon, founded in 2006. It offers compute power, database storage, content delivery, IoT, analytics and big data, identity access, monitoring & management, and other functionalities to help organizations meet their business needs, scale and grow.

3.2. Google Cloud Platform

Google Cloud Platform [6] is a set of cloud computing services, offered by Google, founded in 2008. It offers compute power, database storage, content delivery, IoT, analytics and big data, identity access, monitoring & management, and other functionalities to help organizations meet their business needs, scale and grow.

3.3. Microsoft Azure

Azure [4] is set of cloud services from Microsoft, founded in 2010, which offers compute power, database storage, content delivery, IoT, analytics and big data, identity access, monitoring & management, and other functionalities to help organizations to meet their business needs, scale and grow.

4. Sentiment Analysis (SA) Methodology

Twitter [3] as a communication and social networking service with a quite complex tweet structure which contains a very specific length of text (140 characters), various hashtags, URLs. The limit imposed upon individual posts, causes users to often utilize shorthand notation as well as emoticons in sentiment expression [26]. For this reason, tweets are quite different from others text formats such as: new articles or product reviews. To solve this problem many sentiment

analysis methods have been developed for the tweet's analysis. These methods can be divided into two categories: Distantly supervised methods [27] and fully supervised methods [28].

The main problem with twitter information is how to extract the information that is available on tweets and how it can be used to draw meaningful insights. For this reason, we propose a methodology for tweet sentiment analysis (TSA). For our study, the following steps are proposed as depicted in **Figure 2**.

4.1. Collect Tweets Dataset

Twitter's API allows for complex queries like pulling every tweet about a certain topic. Data extracted from Twitter using Stream and Search functions are available in Twitter's APIs, which allows to do different queries as needed. Certain keywords and hashtags were used in the search process such as "azure, azurecloud, aws, awscloud, amazoncloud, gcpcloud, googlecloud". During the extraction process, we have excluded tweets from cloud provider's official accounts; for instance, tweets from accounts like "Azure, Microsoft, AzureSupport, awscloud, AWS_Partners, Amazon, Google, GoogleCloud, and GcloudPartners" were not collected. The number of tweets collected were 11,822 for the three cloud providers. The distribution of the tweets data is presented in the following **Table 1**.

4.2. Preprocessing Data and Cleaning Tweets

The tweets collected were cleaned from hashtags, URLs, unwanted strings, mentions, digits, none words & spaces and converted the words in all the tweets to lower case. Then, we created a label feature to hold the outcome of the classification process. For the tweet's classification into positive, negative or neural is a challenging task; we have used a dictionary of positive and negative opinion words to classify the tweets; opinion lexicon [29] [30], which contains around 6800 positive and negative words.

Table 1. Tweets dataset distribution.

Cloud Providers	#AWS	#Azure	#Google	Total
No of Tweets	3970	3992	3860	11,822

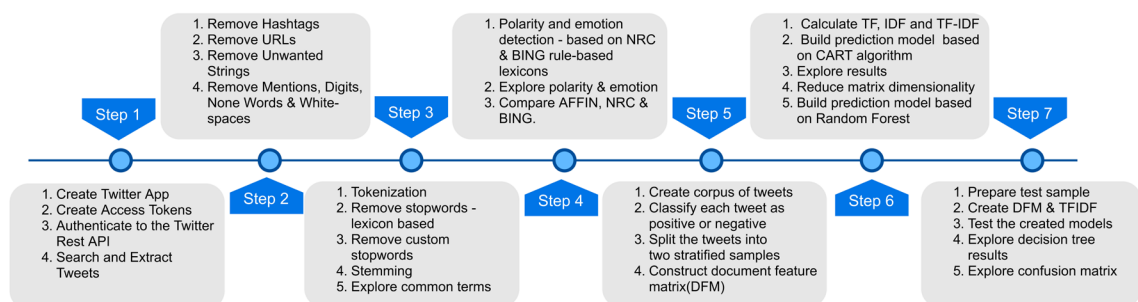


Figure 2. Tweet sentiment analysis methodology (Source: Authors).

4.3. Tokenization of Tweets

In this step, we used the tokenization technique on all the tweets of our dataset as shown in **Figure 3**.

We did find out that, the most frequent words were stop words like “to, the, and, a, of, etc.”; therefore, we have removed them based on the three lexicons “onix, smart and snowball” as shown in **Figure 4**. However, it should be noted that, removing stop words can cause a problem, if the context is affected.

After the stop words removal, some noise appears in the data caused by certain keywords like cloud providers names and some general common terms in cloud computing world.

Now, and as it can be seen in **Figure 5**, there is a clear indication about the overall most frequent words and specific cloud provider’s most frequent words.

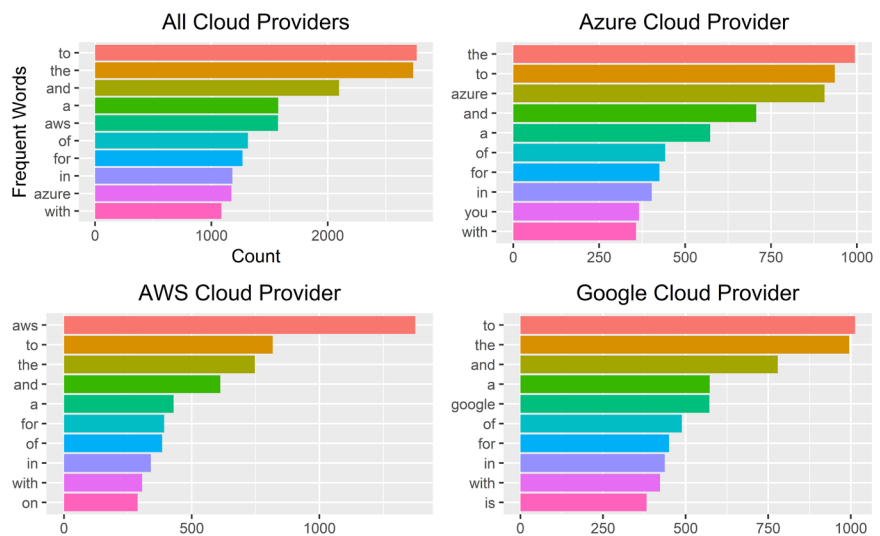


Figure 3. Tokenization of Tweets.

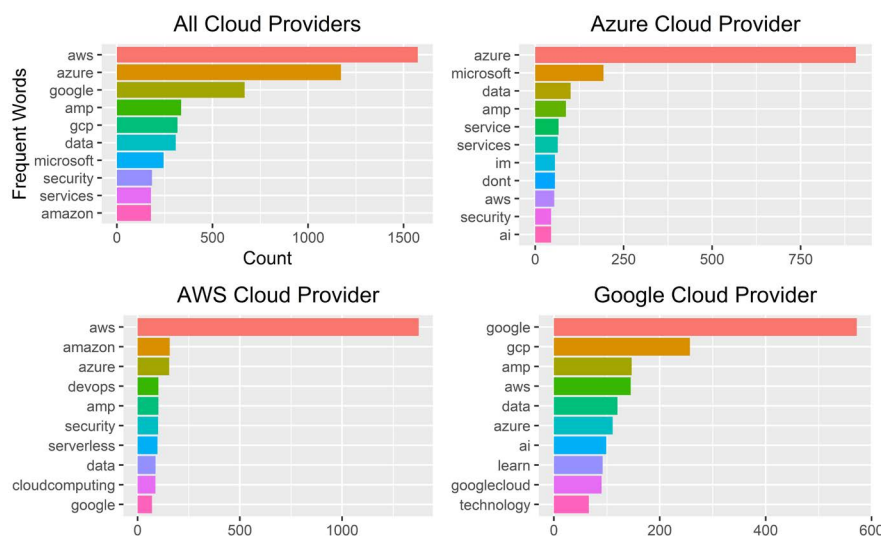


Figure 4. Tokenization of cleaned Tweets.

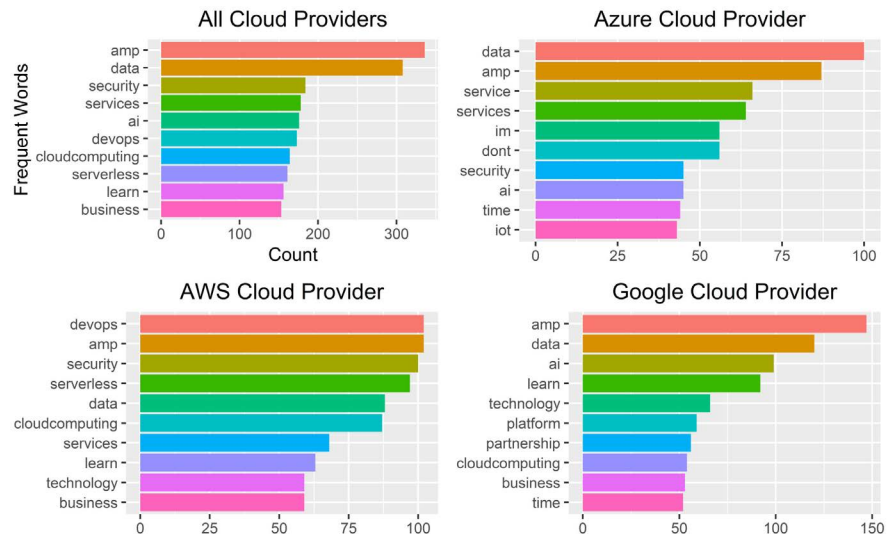


Figure 5. Most frequent words.

4.4. Polarity and Emotion Detection

In this step, the polarity and emotion technique are being evaluated [30] [31]. We can note that, the prominent positive terms in Azure are: easy, improve, happy, love, beautiful. On the other hand, the negative terms are: miss, issue, bad, error. In Google Cloud, the positive prominent terms are: excited, innovation, support, happy, easy etc., and the negative terms are: issue, miss, error, hard etc. In AWS, the positive prominent terms are: innovation, support, easy, excited etc., and the negative terms are miss, threat, idle, error. Since most of the cloud computing services provided are similar across different providers, many positive and negative prominent terms are common in all cloud providers as shown in the following **Figure 6**.

As shown in **Figure 7**, the overall polarity of the tweets shows 68.5% positive and 31.5% negative percentage. The polarity percentage across the cloud providers is close to each other, for instance, Azure shows 63.8% positive and 36.2% negative, Google Cloud shows 72.6% positive, 27.4% negative, and AWS shows 69.1% positive and 30.9% negative.

Looking at the same from a different angle, using emotion classification shows that the positive, emotion and trust are always ahead of other emotions in all cloud providers as presented in **Figure 8**.

Investigating the different emotions about cloud computing services represented by the cloud providers (Azure, Google Cloud and AWS); the most frequent words in the positive emotion are “learn, technology, innovation, excited etc.”. Anticipation emotion’s frequent words are “time, excited, happy, improve” are shown in the following **Figure 9**.

On the other hand, words involved in Anger emotion are “ill, challenge, bad, threat disaster etc.”, whereas, sadness emotion shows high frequency for “ill, error, bad, disaster”. Results are shown in **Figure 9**.

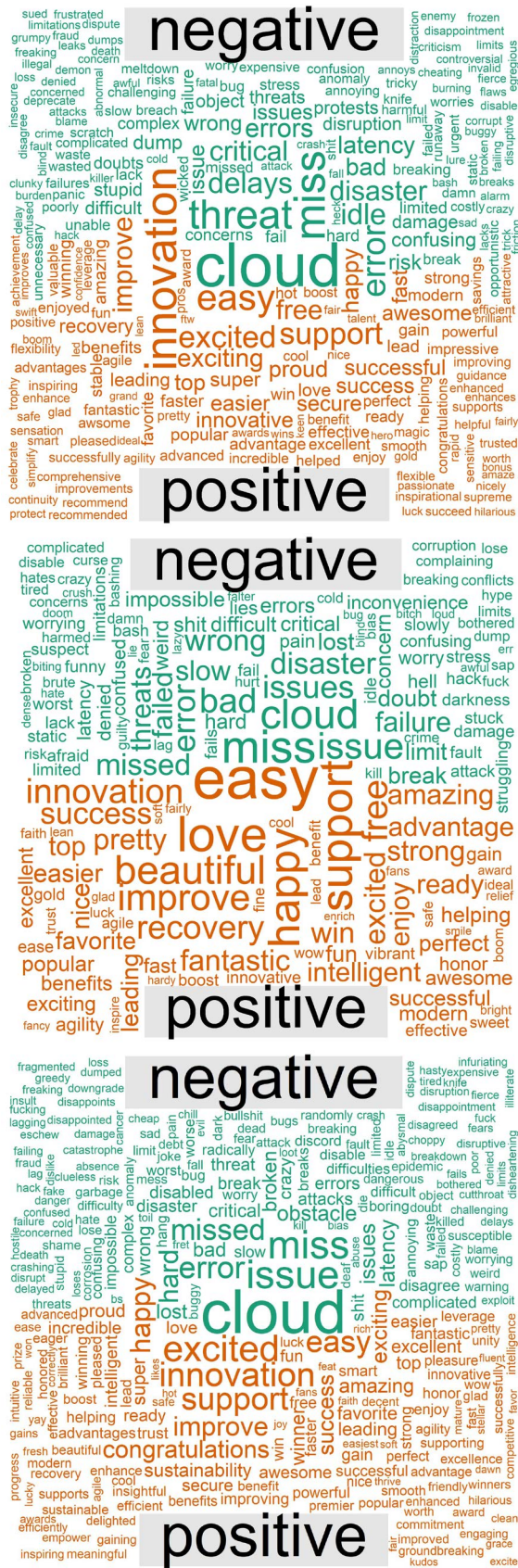


Figure 6. Positive and negative terms (Left: Azure, Center: Google, Right: AWS).

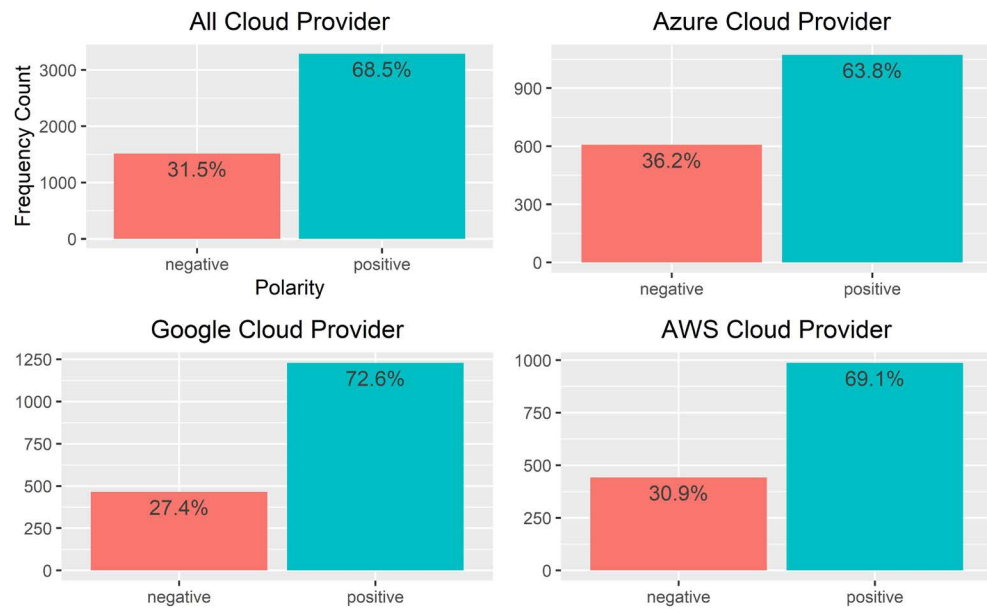


Figure 7. Sentiment analysis.

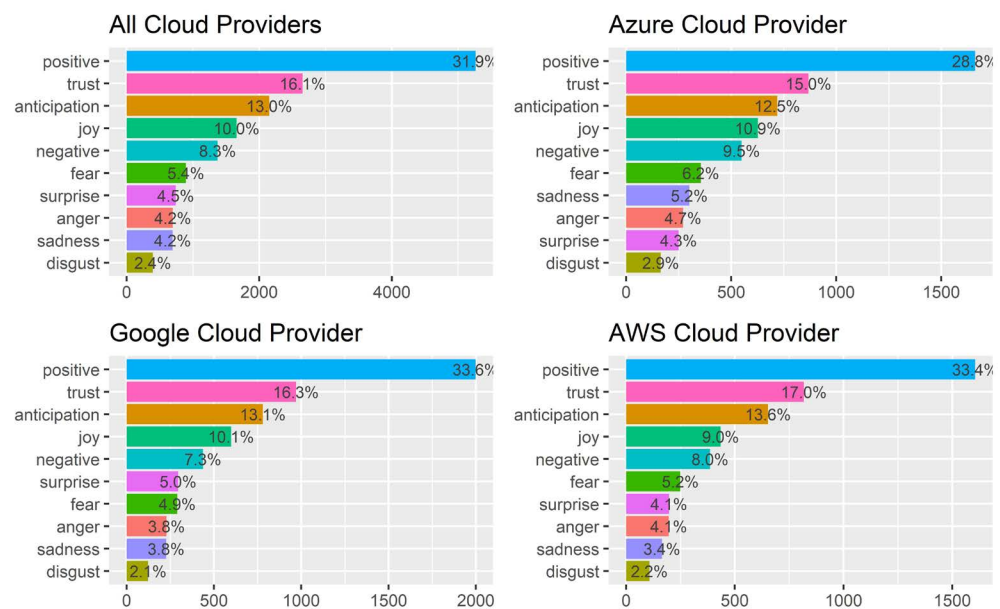


Figure 8. Emotional vocabulary.

Lexicon Comparison

The results of the three lexicons (AFFIN [32], NRC [33], BING [34]) show different sentiments as shown in the following Figure 10; however, the peaks and dips in the sentiment are about the same in the tweets of the cloud providers; however, AFFIN lexicon shows positive sentiments more than the NRC and BING lexicons.

4.5. Build the Document Frequency Matrix (DFM)

The Bag of Words Model [35] iterates through all the tweets in our dataset and

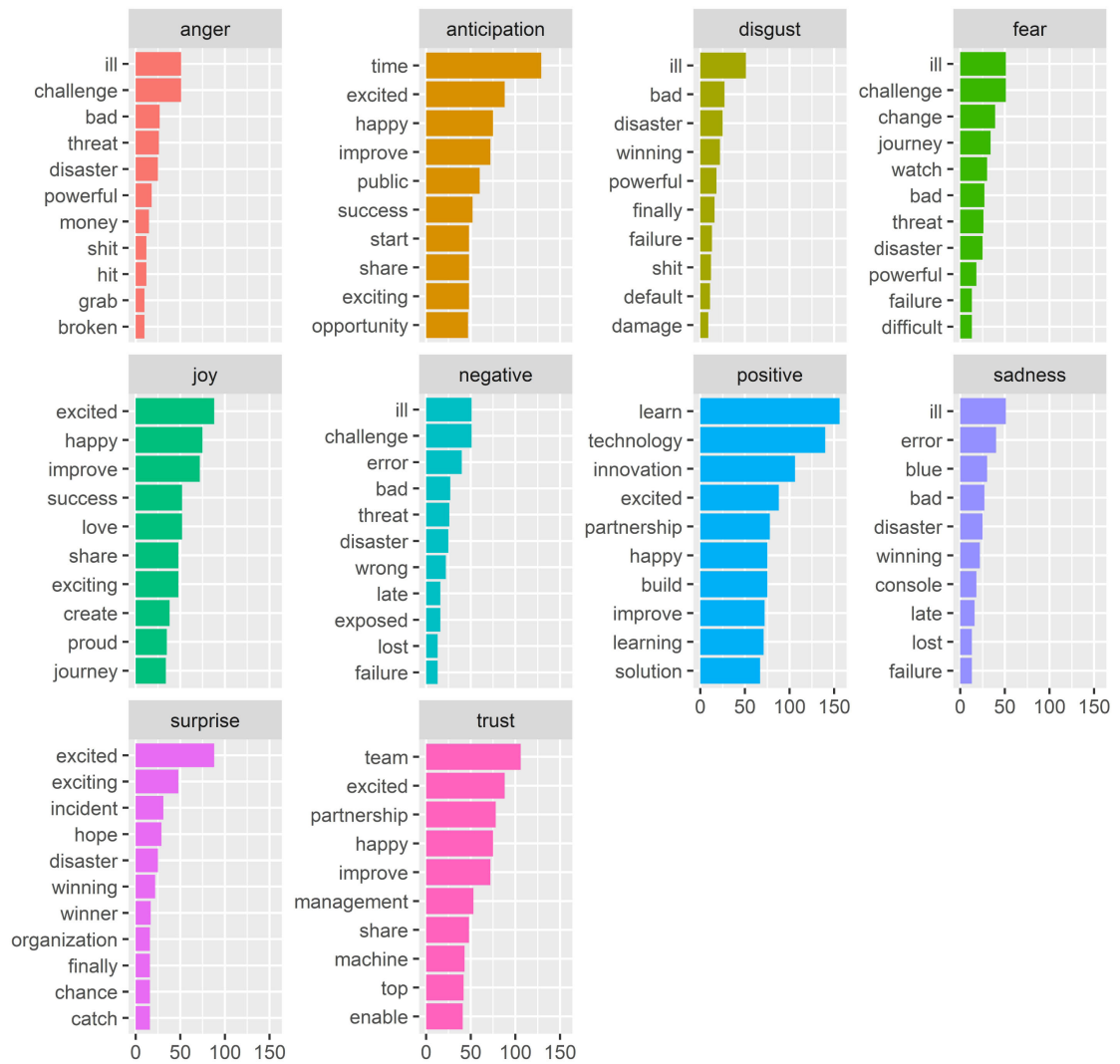


Figure 9. Emotion detection word (Anger, Anticipation, Disgust, Fear, Joy, Negative, Positive; Sadness, Surprise, Trust).

calculates its frequency of occurrence in that dataset. The frequency for each term in each tweet is stored in the Document Frequency matrix (DFM), in order to compare all the tokens or terms across each tweet. In this matrix, each row represents a tweet and each column describes a term or feature in that tweet. A sample of the DFM matrix is presented in **Figure 11**.

The above DFM matrix will help in next steps of analytics classification and performance investigation; however, the document of the frequency matrix (DFM) can become a very large sparse matrix. This depends on the number of tweets in the corpus and the words in each tweet, which will be very difficult and time consuming once we start training the prediction model using machine-learning algorithms. To solve this issue, a new Vector Space Model is introduced using a Latent Semantic Analysis (LSA) technique [35]. The scalability and dimensionality problem are handled by the Vector Space Model. This model represents tweets as vectors of numbers. Given DFM matrix the dot product of the tweets is

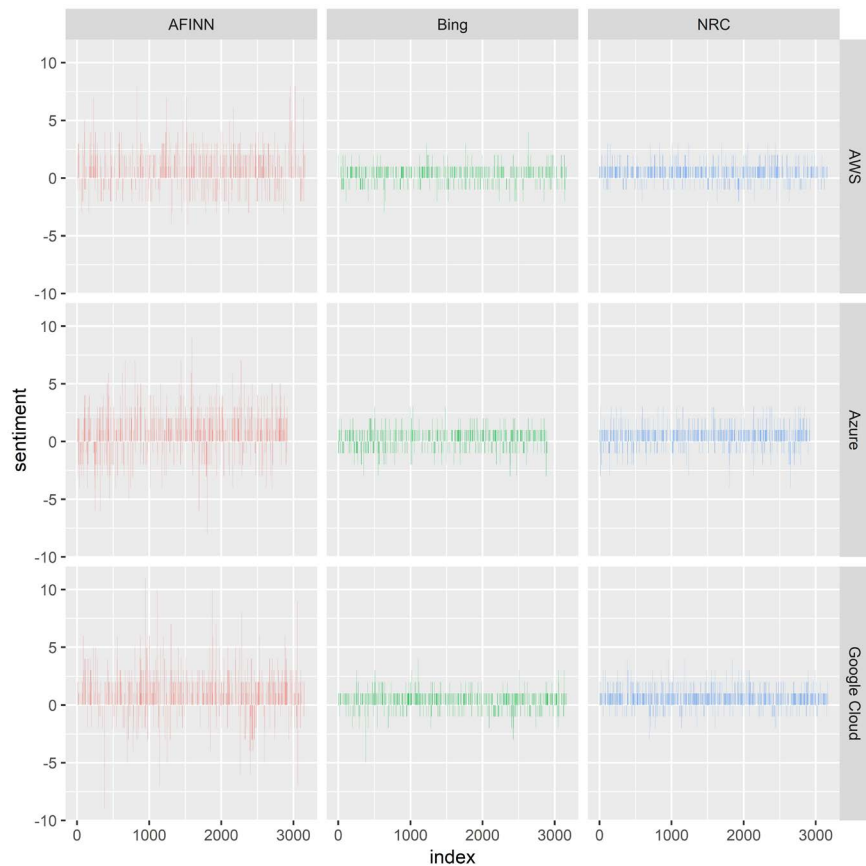


Figure 10. Lexicon comparison.

Document-feature matrix of: 10 documents, 15 features (76.0% sparse).
 10 x 15 sparse Matrix of class "dfm"

docs	features														
text1	2	1	2	1	1	1	1	3	1	1	2	1	1	1	1
text2	3	0	2	1	0	0	0	2	0	0	0	0	0	0	0
text3	2	0	0	2	0	0	0	2	0	0	0	1	0	0	0
text4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
text5	0	0	1	0	0	0	0	3	0	0	0	0	0	0	0
text6	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
text7	1	0	0	1	0	0	1	3	0	0	0	2	0	0	0
text8	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0
text9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
text10	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 11. Document frequency matrix (DFM).

indicative of the document correlation given the set of matrix terms. This paper uses the Latent Semantic Analysis (LSA) technique. This technique is based on extracting the relationship between tweets and terms, assuming that terms which are close in meaning will appear in similar pieces of text. LSA leverages a Singular Value Decomposition (SVD) factorization of term-document matrix to extract these relationships as seen in Equation (1).

$$SVD = V \sum W^T \tag{1}$$

where:

V contains the eigenvectors of the term correlations XX^T .

W contains the eigenvectors of the tweet correlations $X^T X$.

The Latent Semantic Analysis, often provides a solution for the dimensionality problem in text analytics with the selection of the most important singular values. However, the selection of the values of SVD computationally intensive for the prediction models.

Adding a label feature (positive/negative) based on our previous rules-based lexicon's classification. After splitting the tweets into 70% - 30% stratified splits, we proceeded with the machine learning algorithms prediction models.

4.6. Prediction Model Analysis-Classifiers

In order to predict the tweet sentiment, we build the sentiment classifiers using different models such as Decision Tree/CART and Random Forest. An important criterion for a good classification procedure is, not only the production of accurate classifiers but also, the provision of insights and understanding into the predictive structure of the data [10]. We have tested our classifiers on a set of Twitter posts hand-annotated. A brief description of the two classification models are presented in the following sections.

4.6.1. Decision Tree-CART

Decision trees are supervised learning models. In this analysis, we used the classification and regression trees (CART) [36] to predict the outcome; variable positive or negative (Tweets). Using CART as the base, this classifier extracts knowledge from the proposed dataset and then classifies effectively new data. An example of this CART tree is presented in the following Figure 12.

4.6.2. Random Forest

A random forest [36], is a classifier consisting of a collection of tree structured classifiers $\{h(x, \Theta_k), k = 1, \dots\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors, and each tree casts a unit vote for the most popular

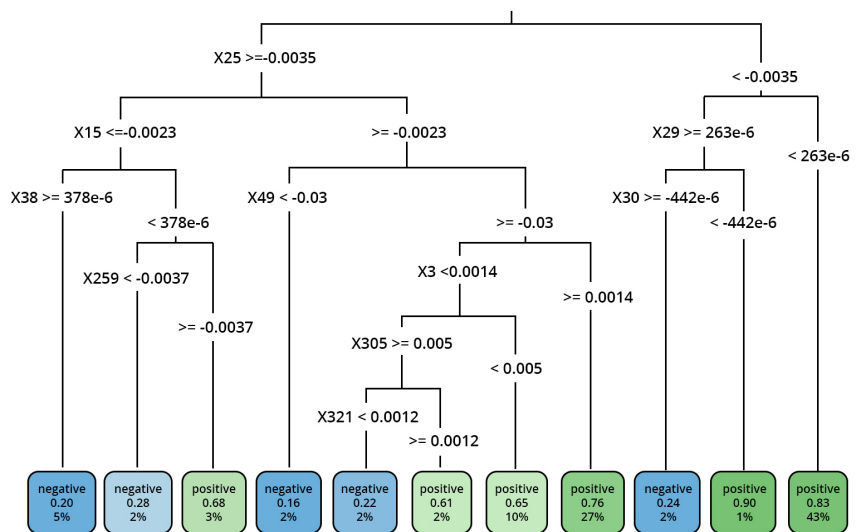


Figure 12. CART classifier.

class at input x . Random forests are powerful algorithms, which operate by constructing multiple decision trees; therefore, it is costly in terms of resources and time is required for training. Since we have already reduced our matrix dimension using SVD, it is possible to use the random forest algorithm. In addition to the usage of random forest, a feature added to the matrix to calculate the mean cosine similarity for all the negative tweets, assuming low cosine similarity for negative tweets.

As shown in **Figure 13**, the use TF-IDF (Term Frequency-Inverse Document Frequency) to enhance the observation, the model and the accuracy is 0.775 at the value $cp = 0.01825293$.

5. Classifiers and Performance Evaluation

For the performance evaluation, the two main factors are the accuracy and the sensitivity using a confusion matrix. Accuracy is calculated as the percentage of the total number of instances present for classification which were correctly classified by the model developed. In this paper, accuracy is measured on a scale of 0 to 100 and not from 0 to 1. The confusion matrix [36] presents the number of either type correctly and incorrectly classified by the model developed.

Based on the below **Table 2** the accuracy is defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} = \frac{a + d}{a + b + c + d} \quad (2)$$

The sensitivity is referred as the true positive (recognition rate) and it is calculated as the percentage of True Positive (TP) over the sum of TP and FN .

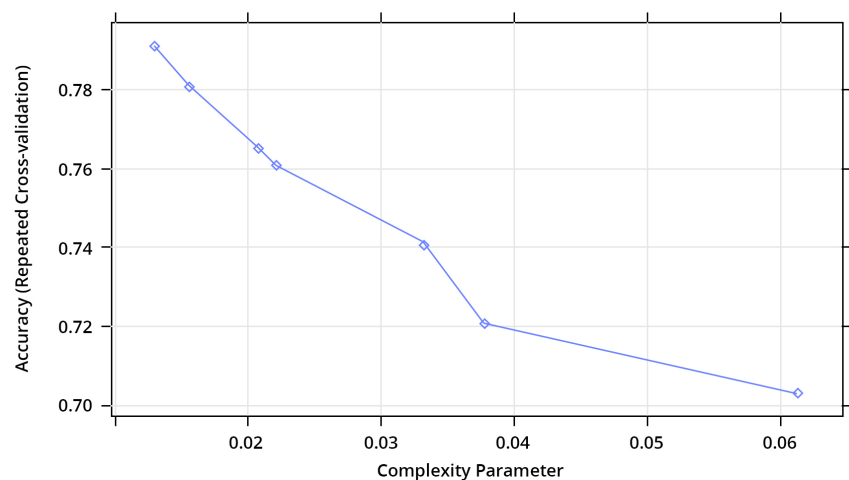


Figure 13. Accuracy using TF-IDF approach.

Table 2. Confusion matrix for the evaluation of classification models [36].

		Predicted Class	
		Yes	No
Actual Class	Yes	a (TP)	b (FN)
	No	c (FP)	d (TN)

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{a}{a + b} \quad (3)$$

The specificity is referred as the true negative rate and it is calculated as the percentage of True Negative (TN) over the sum of FP and TN .

$$\text{Specificity} = \frac{TN}{FP + TN} = \frac{d}{c + d} \quad (4)$$

5.1. Experiments and Results

For all the scenarios, a k -fold cross-validation method has been employed to divide the dataset into different sub-samples. A k -fold cross-validation method, is a way to improve the holdout method. The original sample is randomly divided into k equally (or nearly equally) sized sub-samples, and the cross-validation process is repeated k times (the folds).

Each time, one of the k sub-samples is used as the test set and the other $(k - 1)$ sub-samples are put together to form the training set. Finally, the average error across all k trials is computed [36]. For our analysis, we explored the computed results based on 10 folds. Cross-validation was repeated 3 times over the training portion of our dataset that we converted previously to document term matrix in combination with the various classifiers.

In greater details, the Bag of Words Model is applicable on the train dataset. The Data Frequency Matrix (DFM) stores the token which in turn are converted to a new Train Dataset. To the first model, a 10-fold cross validation is applied (three times on the Train data set) and then the CART Decision Tree Classification Algorithm as described in earlier sections of this paper. In question to efficiency increase, the TF-IDF model is applied and associated classification is based again on 10-fold Cross Validation (again three times), followed by the CART Decision Tree classification technique.

Last, in order to improve this efficiency reduction, the Vector Space Model (Latent Semantic Analysis and Singular Value Decomposition) are applied on the dataset for both classification algorithms (CART and RANDOM FOREST).

Table 3, presents a summarized view of the operations carried out on the dataset, along with the parameters used to measure the efficiency of the model developed.

From the above tabulated results, the LSA-SVD & CART model predicts 565 true negative and 1612 true positive (see **Table 4**). The accuracy measurement has been improved using the random forest algorithm with the SVD and cosine

Table 3. Performance evaluation using confusion matrix.

Models used	Accuracy	Sensitivity	Specificity	True Positive	True Negative
LSA-SVD & CART	80.35%	68.50%	85.28%	649	226
LSA-SVD & RANDOM FOREST	85.61%	77.5%	88.86%	1612	565

and migration. This paper discussed the cloud computing and services in terms of customers' opinions, feedbacks, attitudes, and emotions by using sentiment analysis. The aim is to help people and organizations to understand the benefits and challenges of cloud services from the general public's perspective view as well as opinions about existing cloud providers. Focus was given on three main cloud providers: Azure, Amazon Web Services (AWS) and Google Cloud.

The methodology used in this paper is based on sentiment analysis applied to tweets extracted from social media platform (Twitter) via its search API. We have extracted a sample of 11,000 tweets and each cloud provider has almost similar proportion of the tweets based on relevant hashtags and keywords. The extracted data, exclude tweets from the official cloud provider's accounts.

Bing and NRC Lexicons were used to measure the polarity and emotion of the tweets' terms. Results indicated that the overall polarity classification, across all cloud providers showed 68.5% positive and 31.5% negative. More specifically, Azure showed 63.8% positive and 36.2% negative, Google Cloud showed 72.6% positive and 27.4% negative and AWS showed 69.1% positive and 30.9% negative.

6.2. Future Work

In the near future we are aiming to measure the associated performance using SVM (Support Vector Machine), Neural Networks in supplement to the methods used and described in this paper. Moreover, to be in a position to have more clear idea on the analysis of the given sentiment, we are planning to proceed with a more detailed analysis in order to investigate which one suits better.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Cisco (2018) Cisco Global Cloud Index: Forecast and Methodology, 2016-2021. White Paper.
<https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html>
- [2] Horrigan, J.B. (2008) Online Shopping. Pew Internet & American Life Project Report, Washington.
- [3] Twitter <https://www.twitter.com>
- [4] Azure <https://www.azure.com>
- [5] AWS <https://www.aws.amazon.com>
- [6] Google <https://www.google.com>
- [7] Bricker, E. (2011) Can Social Media Measure Customer Satisfaction? Net Base, Berkeley, CA.
- [8] Hill, N., Brierley, J. and MacDougall, R. (2017) The Business Case for Customer Satisfaction. In: *How to Measure Customer Satisfaction*, Taylor & Francis Group,

- London, 118-121. <https://doi.org/10.4324/9781315253107-14>
- [9] Faizaan, M., Nawaz, M., Khan, S., Shaukat, Z. and Aslam, N. (2011) Impact of Customer Satisfaction on Customer Loyalty and Intentions to Switch. *International Journal of Business and Social Science*, **2**, 263-270.
- [10] Kavis, M. (2014) *Architecting the Cloud*. Wiley, New York.
<https://doi.org/10.1002/9781118691779>
- [11] Izumi, K., Suzuki, H. and Toriumi, F. (2017) Transfer Entropy Analysis of Information Flow in a Stock Market. In: Aruka, Y. and Kirman, A., Eds., *Economic Foundations for Social Complexity Science: Theory, Sentiments and Empirical Laws*, Springer, Berlin, 193-208. https://doi.org/10.1007/978-981-10-5705-2_10
- [12] Jegadeesh, N. and Wu, D. (2015) Deciphering FedSpeak: The Information Content of FOMC Meetings. 2016 *AFA Annual Meeting Working Paper*. San Francisco Marriott Marquis, San Francisco, CA, 3-5 January 2016.
<https://doi.org/10.2139/ssrn.2939937>
<https://www.aeaweb.org/conference/2016/retrieve.php?pdfid=1136>
- [13] Hamouda, A., Marei, M. and Rohaim, M. (2011) Building Machine Learning Based Senti-Word Lexicon for Sentiment Analysis. *Journal of Advances in Information Technology*, **2**, 199-203. <https://doi.org/10.4304/jait.2.4.199-203>
- [14] Zhang, Z.H. and Lan, M. (2015) Learning Sentiment-Inherent Word Embedding for Word-Level and Sentence-Level Sentiment Analysis. 2015 *International Conference on Asian Language Processing*, Suzhou, 24-25 October 2015, 94-97.
<https://doi.org/10.1109/IALP.2015.7451540>
- [15] Pang, B., Lee, L. and Vaithyanathan, S. (2002) Thumbs Up? Sentiment Classification Using Machine Learning Techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Philadelphia, PA, July 2002, 79-86.
<https://doi.org/10.3115/1118693.1118704>
- [16] Turney, P. (2002) Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Philadelphia, PA, 7-12 July 2002, 417-424. <https://doi.org/10.3115/1073083.1073153>
- [17] Pfitzner, R., Garas, A. and Schweitzer, F. (2012) Emotional Divergence Influences Information Spreading in Twitter. ICWSM-12.
- [18] Garas, A., Garcia, D., Skowron, M. and Schweitzer, F. (2012) Emotional Persistence in Online Chatting Communities. *Scientific Reports*, **2**, Article No. 402.
<https://doi.org/10.1038/srep00402>
- [19] Alexander, P. and Paroubek, P. (2010) Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *LREC2010*, Valletta, Malta, 17-23 May 2010.
- [20] Go, A., Bhayani, R. and Huang, L. (2009) Twitter Sentiment Classification Using Distant Supervision. CS224N Project Report, 1-12.
- [21] Schmid, H. (1994) Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, **12**, 1-9.
- [22] Aston, N., Liddle, J. and Hu, W. (2014) Twitter Sentiment in Date Streams with Perceptron. *Journal of Computer and Communications*, **2**, 11-16.
<https://doi.org/10.4236/jcc.2014.23002>
- [23] Peter, M. and Timothy, G. (2011) The NIST Definition of Cloud Computing: Recommendations of the National Institute of Standards and Technology.
<https://doi.org/10.6028/NIST.SP.800-145>

- <https://csrc.nist.gov/publications/detail/sp/800-145/final>
- [24] Mell, P.M. and Grance, T. (2011) The NIST Definition of Cloud Computing. <https://doi.org/10.6028/NIST.SP.800-145>
- [25] Karamitsos, I. and Apostolopoulos, C. (2013) Convergence: Smart Home into Cloud. *10th IEEE International Conference Wireless and Optical Communications Networks (WOCN2013)*, 26-28 July 2013, Bhopal, India, 1-4. <https://doi.org/10.1109/WOCN.2013.6616254>
- [26] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D. and Kappas, A. (2010) Sentiment Strength Detection in Short Informal Text. *Journal of the American Society for Information Science and Technology*, **61**, 2544-2558. <https://doi.org/10.1002/asi.21416>
- [27] Go, A., Bhayani, R. and Huang, L. (2009) Twitter Sentiment Classification Using Distant Supervision. Technical Report.
- [28] Jansen, B.J., Zhang, M., Sobel, K. and Chowdury, A. (2009) Twitter Power: Tweets as Electronic Word of Mouth. *Journal of the American Society for Information Science and Technology*, **60**, 2169-2188. <https://doi.org/10.1002/asi.21149>
- [29] Qaisi, L.M. and Aljarah, I. (2016) A Twitter Sentiment Analysis for Cloud Providers: A Case Study of Azure vs. AWS. *2016 7th International Conference on Computer Science and Information Technology*, Amman, Jordan, 13-14 July 2016, 1-6. <https://doi.org/10.1109/CSIT.2016.7549473>
- [30] Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T. and Qin, B. (2014) Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, **1**, 1555-1565. <https://doi.org/10.3115/v1/P14-1146>
- [31] Trindade, L., Wang, H., Blackburn, W. and Rooney, N. (2013) Effective Sentiment Classification Based on Words and Word Senses. *2013 International Conference on Machine Learning and Cybernetics*, Tianjin, 14-17 July 2013, 277-284. <https://doi.org/10.1109/ICMLC.2013.6890481>
- [32] Pham, D.-H., Le, A.-C. and Le, T.-K.-C. (2018) Learning Word Embeddings for Aspect-Based Sentiment Analysis. In: Hasida, K. and Pa, W., Eds., *Computational Linguistics. Communications in Computer and Information Science*, Springer, Singapore, 28-40. https://doi.org/10.1007/978-981-10-8438-6_3
- [33] Khan, F.R., Al-balushi, H.Y., Algaithi, A.D. and Al-Shihi, A.A. (2017) Special issue on ISSN-2456-205X Economic Growth in Oman amidst Oil Crisis. *Ahead International Journal of Recent Research Review*, **1**, 154.
- [34] Agarwal, B. and Mittal, N. (2015) Machine Learning Approach for Sentiment Analysis. In: *Prominent Feature Extraction for Sentiment Analysis. Socio-Affective Computing*, Springer, Cham, 21-45. https://doi.org/10.1007/978-3-319-25343-5_3
- [35] Aggarwal, C.C. (2018) Opinion Mining and Sentiment Analysis. In: *Machine Learning for Text*, Springer, Cham, 413-434. https://doi.org/10.1007/978-3-319-73531-3_13
- [36] Han, J., Kamber, M. and Pei, J. (2012) *Data Mining Concepts and Techniques*. 3rd Edition, Morgan Kaufmann Publishers, San Francisco, CA.