Sentiment Analysis Algorithms through Azure Machine Learning: Analysis and Comparison

Osama Harfoushi[1], Dana Hasan[2] & Ruba Obiedat[3]

[1] Department of Business Information Technology, The University of Jordan, Jordan

[2] Microsoft Dynamics technical Consultant, ebecs A DXC Technology Company, Jordan

[3] Department of Business Information Technology, The University of Jordan, Jordan

Correspondence: Osama Harfoushi, Department of Business Information Technology, The University of Jordan, Jordan. Tel: 962-795-707-656. E-mail: o.harfoushi@ju.edu.jo

Abstract

The Sentimental Analysis (SA) is a widely known and used technique in the natural language processing realm. It is often used in determining the sentiment of a text. It can be used to perform social media analytics. This study sought to compare two algorithms; Logistic Regression, and Support Vector Machine (SVM) using Microsoft Azure Machine Learning. This was demonstrated by performing a series of experiments on three Twitter datasets (TD). Accordingly, data was sourced from Twitter a microblogging platform. Data were obtained in the form of individuals' opinions, image, views, and twits from Twitter. Azure cloud-based sentiment analytics models were created based on the two algorithms. This work was extended with more in-depth analysis from another Master research conducted lately. Results confirmed that Microsoft Azure ML platform can be used to build effective SA models that can be used to perform data analytics.

Keywords: sentiment analysis, machine learning, Microsoft azure, logistic regression algorithm, support vector machine algorithm

1. Introduction

The Sentimental Analysis is a widely known and used technique in the natural language processing realm. It is often used in determining the sentiment of a text. It encompasses studying peoples' attitudes, feelings and opinions towards a product, an event or organization computationally (Kasture & Bhilare, 2017; Li & Wu, 2010; Thomas, et al., 2011). It can be used to assess reviews posted by people online about their decisions regarding the food they consume, items the use and other issues affecting them. As such, sentiment analysis involves assessing a piece of writing intending to determine whether is neutral, negative or positive. It is often applied in several areas namely plagiarism checking; intellectual property, social media analytics, product reviews, and document/case classification. In social media analytics, which is the focus on the present study, studies have demonstrated the possibility of using sentiment analysis (SA) through platforms like Microsoft Azure Machine learning; Amazon SageMaker and Amazon Machine Learning; and Google Cloud Machine Learning to analyse social media analytics.

For example, Liu, et al. (2015) used Microsoft Azure Machine learning to perform Twitter sentiment analysis and to develop a model for classifying machine learning that allows for the identification of tweet sentiments and content that illustrate positive-value user contribution. Liu, et al. (2015) used Al-powered cognitive and data mining tools to analyse factors of social influence. The predictive sentiment analysis model developed from this study encompassed a combination of custom-developed natural language model and a traditional supervised machine language algorithm for identifying promotional tweets. In a similar study, Qaisi & Aljarah (2016) used sentiment analysis performed through Microsoft Azure and Amazon machine learning to analyze the opinions and reviews of Amazon and Microsoft. Results confirmed the possibility of using sentiment analysis via Azure and Microsoft to perform social media analytics. It was revealed based on the sentiment analysis that Azure had more positive tweets (65%) than Amazon (45%) and that Amazon had more negative polarity (50%) than the Microsoft Azure (25%). Similarly, Barbosa & Feng (2010) used sentiment analysis to classify data obtained from Twitter and proposed that syntax features such as links, exclamation marks, punctuation, retweets, and hashtags should be

used alongside POS of words and polarity in performing sentiment analysis.

Results of these studies demonstrate that the sentiment analysis built on various machine language platform namely Microsoft Azure or any other platform can be used to perform data analytics. However, there is hardly studies that have used demonstrated the side by side use of sentiment analysis built on Azure ML for social media analytics. This study demonstrated that Microsoft Azure Machine Learning (ML) based on two Machine Learning (ML) algorithms: Logistic Regression, and Support Vector Machine (SVM) can be used to build sentiment analysis (SA) models used to perform data analytics. Moreover, a comparison between the two algorithms is carried out.

2. Literature Review

2.1 Microsoft Azure Machine Learning

Microsoft Azure Machine Learning encompasses cloud services that enable the creation, deployment, and management of applications by developers via a global network of datacentres for Microsoft. This cloud computing model emphasizes the cloud platform's differentiating features namely flexibility, agility and scalability. Currently, Azure calculates the contribution score of the user based on social media metrics. This allows for the easy quantification of the value of users of Microsoft add to its cloud business on social media to enable it to provide differentiated services.

Azure ML also supports multiple ML algorithms related to regression, classification, and clustering. It allows for the customization of models using python and R (Qasem et al., 2015). Azure ML studio allows for the dragging and dropping of Modules and datasets (i.e., Ml algorithms, feature selection, and pre-processing) and links them together. This experiment can be trained and transformed into a predictive experiment. This predictive experiment allows users to build their models (Ericson et al., 2016; Rajpurohit, 2014).

In general, Microsoft Azure is designed to set a playground for experienced and newcomers data scientists. It provides a variety of algorithms with only a single clustering algorithm. Azure ML is often characterized by the Cortana Intelligence Gallery, which is a collection of ML solutions created by the community to be reused and explored by data scientists. Azure services can be categorized into two: Azure Bot Service and Azure Machine Learning Studio.

Azure ML studio requires users to complete all the operations manually. This includes, data preprocessing, exploration, validating modeling results, and choosing methods. It supports about 100 techniques that address regression, anomaly detection, classification (binary and multiclass), text analysis, and recommendation.

2.2 Machine Learning Algorithms

There are different machine language algorithms such as Support Vector Machine (SVM), Logistic Regression, Network Regression (NNR), and Decision Forest. Logistic regression is a statistical linear algorithm used in task classification. It is usually used to solve simple problems. It can be used as a prediction model. It predicts values by applying statistical analysis (Chen, 2011). The Support Vector Machine algorithm is supervised learning approach used to solve classification problems. It accepts labelled training data and produces hyperplane which is used to maximize the margin between high-dimensional space classes (Wu et al., 2014). The Decision Forest algorithm is a learning method consisting of multiple classification methods. It can construct decision trees each with a different classification. It can perform aggregation and sum histograms to obtain each label's probabilities. The decision forest selects the decision tree with the most votes (Topouzelis & Psyllos, 2012). Neutral Network Regression Algorithm builds a classification model by combining two algorithms: Neural Network and Logistic Regression. It utilizes a logistic function. As such, its output is similar to that of Logistic Regression. It requires the use of a dataset to test an algorithm.

2.3 Sentiment Classification Techniques

There are two approaches to performing SA: lexicon-based approach and ML approach (Devika et al., 2016). ML approach, which is the focus of the present study, is dependent on the training dataset as it involves training the algorithm using a training dataset followed by applying the algorithm to the actual dataset. The classification of SA using the ML approach involves two datasets: testing and training datasets. The classification algorithm utilizes these datasets to verify algorithm performance and to learn dataset. In particular, the training dataset is used in learning dataset while the testing dataset is used in verifying the performance of the algorithm (Sharef et al., 2016).

There are two ways through which ML approaches sentiment classification: supervised learning method; and unsupervised learning method. The supervised learning method utilizes training dataset which includes the score and input label. It enables the classification model to learn using classification algorithms. It is also used in predicting the value for new inputs. On the other hand, the unsupervised learning model does not utilize labelled

dataset. It is trained using datasets involving a group of inputs (Sharef et al., 2016; Tramer et al., 2016).

3. Test Methodology

This study sought to demonstrate that Microsoft Azure ML-based on two Machine Learning algorithms can be used to build sentiment analysis (SA) models used to perform data analytics. This study was extended from a Master thesis with more depth analysis for the data and a new series of experiments to compare the two specific algorithms based on one single machine Learning platform (Hasan, 2017). Moreover, a comparison between the two algorithms' outputs is carried out. This was undertaken using experimental research design. Accordingly, data was sourced from Twitter a microblogging platform. Data were obtained in the form of individuals' opinions, image, views, and twits from Twitter.

Procedurally, SA models were built on the Azure ML platform based on Logistic Regression and Support Vector Machine. Next, the accuracy and performance of these SA models were evaluated. The outcome informed the decision made regarding the machine language that offered the best SA in terms of performance and accuracy. Several experiments were performed and SA model was tested using datasets A, B, and C, and the model executed using each data set.

3.1 Azure Sentiment Analysis Model

Azure SA Model was created on Microsoft Azure. It was used to determine the tweets' sentiment. This was done by building the Azure ML model, training it on how to detect the sentiment, and finally setting it as a predictive model to facilitate it to detect and identify sentiments as neutral, negative or positive.

The sentiment analytics model was created based on the Logistic Regression and Support Vector Machine algorithms. The model was trained using dataset (TD). This was done after subjecting the dataset to normalization, which involved getting rid of punctuations, numbers and stop words as well as removing URLs and emails from the tweets. A Hashing of bit size 10 and with n-grams was also applied to the tweets before training the model using TD. Lastly, modifications were made to the model to make it detect sentiment.

This study utilized Coachella 2015 Twitter sentiment dataset. This dataset was created by "CrowdFlower" data mining company. It has tweets on Coachella arts and music festival which was held in 2015. The classification results are generated be processed connected elements (Vallejos & Mckinnon, 2013).

The original Coachella dataset consisted of 10 columns and 3800 tweets (Figure 1). The columns consisted of 10 associated fields: tweet created; Coachella yn; Coachella sentiment; name; text; retweet count; tweet Id; user time zone; and tweet location. Two columns were used to test Sentiment Analysis algorithm. Coachella sentiment is the first column and encompasses the sentiment of a tweet. The second column is a text of a tweet.

coachella_ sentiment	coachella _yn	name	retweet_ count	text	tweet_ coord	tweet_created	tweet _id	tweet_location	user_time zone
positive	yes	mick eydi tch	0	So HAPPY to announce that me and @NicolaBo tt have twins!!!! (tickets) #Coachell a2015	[- 33.87 10555 8, 151.2 03347 92]	01/07/2015 14:17	5529 5200 0000 0000 00	Sydney	London

Figure 1. Tweet example

In the present study, two columns out of the ten columns were used to test the SA algorithm. The first column was Coachella sentiment that contained the sentiment of a tweet. The values of the sentiment in the Coachella was ascertained through Sentiment Analysis.

This investigation utilized 1250 records out of the 3800 records since others records contained duplicate tweets. These records were categorized four unequal datasets: Datasets A, B and C. Dataset A included a single column with 100 records. It was used to perform the first test. Database B included a single column and had 300 records. It was used to undertake the second test. Dataset C also consisted of a single column and consisted of 500 records. It was used to perform the third test.

3.2 Azure Sentiment Analysis Training Model Process for Coachella

Azure-based Sentiment Analysis training model had several steps used to build Twitter SA for Coachella model (Figure 2). First of all, Coachella DB was created and served as the tagged dataset. It included a label column and tweet text column. It consisted of a total of 350 tweets with 128 tweets representing neutral sentiment, 106 tweets representing negative sentiment, and 116 tweets representing positive sentiment. Dataset file Coachella served as a comma-separated value file. Other options included Azure ML used to import data. This Machine Learning employed several techniques: Azure table; Azure SQL database; Azure blob; and HTML. Data used in this study was obtained from the Coachella comma-separated value file.

With regard to the pre-process text, Coachella DB was the input step. It involved applying pre-processed procedures to each Coachella DB tweet with stop words and numbers being extracted and URLs and e-mail address removed. Verb contractions were also expanded, and duplicate characters eliminated.

Second Step is executing R-script; The modified dataset version of the previous step was the input for this execute R-script step. This stage involves the execution of the R-script. Punctuation is replaced, and special characters and digits are replaced with space and tweet converted into lowercase.

The execution of R-script step was modified and used an input to edit metadata step. This step handled the modified metadata, and was designed for use to alter the definition of text column of the tweet to convert it into a non-categorical format.

The previous step's dataset served as the input to the feature hashing step. The Feature Hashing module in Azure ML is designed in accordance with the Vowpal Wabbit framework (Qasem et al., 2015). This framework is used to hash features into in-memory indexes.

The next step is to rank featured based on Chi-Squared feature selection. A database that encompassed finer features defined by high predictive power was the input of the chi-squared module. High score features were included whereas low-score features were removed.

The multiclass classification model was built from the Logistic Regression algorithm. To train the classification model, the tagged dataset and ML algorithm were provided as inputs to the model (Figure 2). This enabled the trained model to predict the sentiment for new tweets.

The train model was designed to provide the classification model with a trained dataset aimed at discovering patterns. The two inputs to this module were Logistic Regression algorithm model or configured ML model and the trained dataset. The trained model was the outcome. It was used to create the predictive model used in detecting sentiments of tweets regarding the Coachella event.

The score model was used to predict the trained model. It has two appearances (Figure 2). These appearances had trained dataset and trained models as input, and a set-aside dataset for model testing. This score model generated predicted values and the probability of the values that were predicted. The scored dataset was its output and was in performance evaluation.
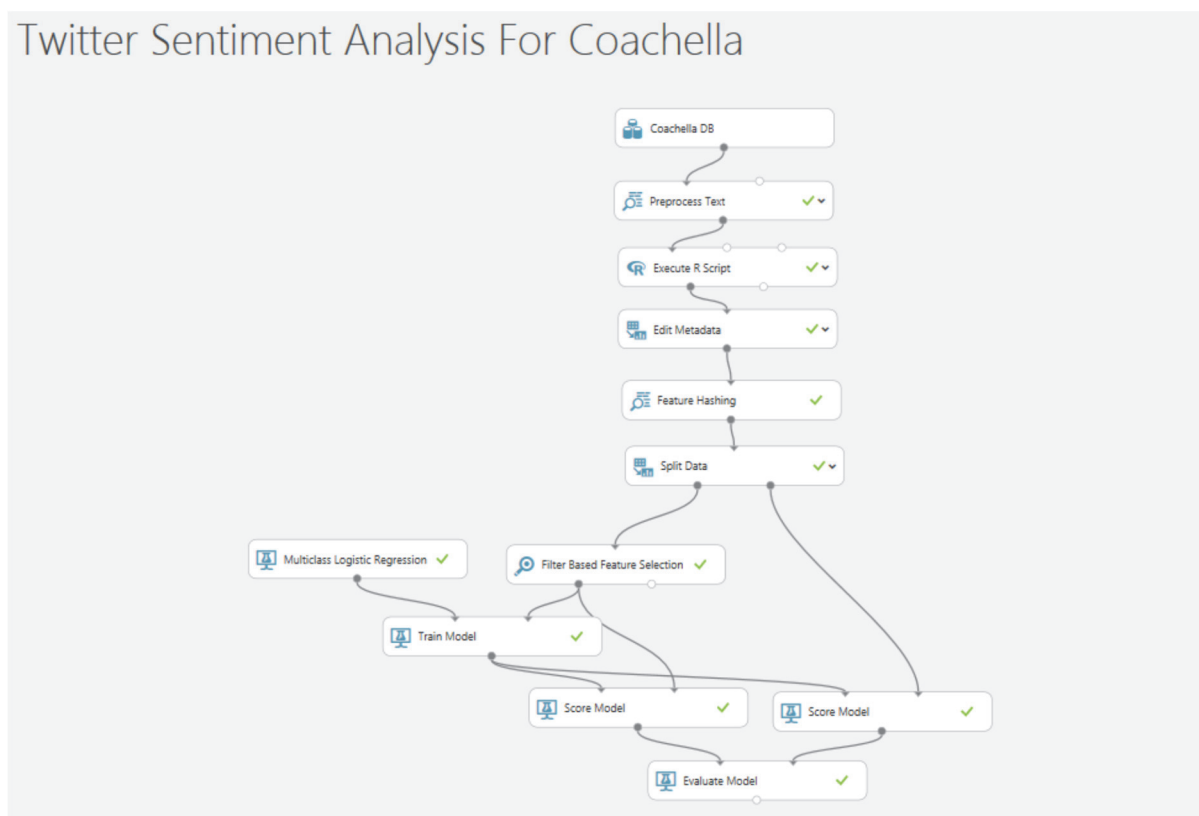
Figure 2. Twitter SA for the Coachella Training Model Process in the Azure ML studio

Finally, the evaluation step measured the trained model's accuracy. Scored dataset and scored trained dataset were the inputs to this step. This model provided the metrics scores used in evaluating Coachella classification model. It also provided an estimation of accuracy for the trained model. It compares the trained model accuracy to data sample of the previous step. Accurate measurements of the classification model, including Recall, Precision, and Accuracy were the output.

3.3 Azure SA Predictive Model for Coachella

The training experiment was converted into the predictive experiment used for sentiment prediction. This predictive experiment was deployed as Azure web service to enable it to receive users' inputs as shown in Figure 3.
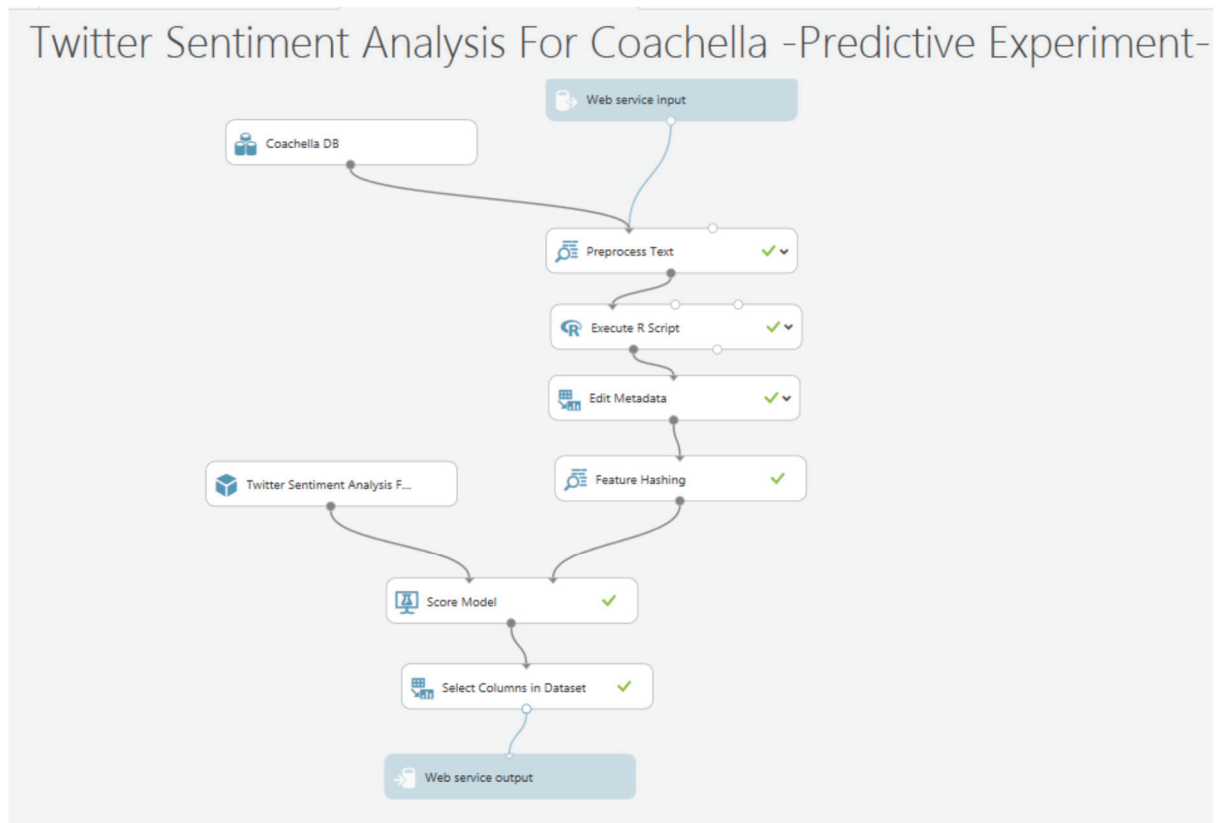
Figure 3. Twitter SA for Coachella Predictive Model Process in Azure ML studio

4. Test Results

The evaluation metrics are used in this experiment in order to assess the quality of SA model that have been created in Azure ML. Usually, Accuracy and Precisions metrics that are used for evaluation of text classification tasks in this test. Accuracy measures how much the algorithm an accurate prediction of the results. Precision measures how values are close to each other. For evaluating the performance of SA model that classifies the tweet as positive, neutral or negative, the confusion matrix is used. Figure 4 represents the true sentiment values (TPOS, TNRU, TNEG) and false sentiment values (FPOS, FNRU, FNEG). Confusion Matrix is used to compare the model predicted values with the actual values.

$$\text{Accuracy} = (TPOS + TNEU + TNEG) \, / \, (TPOS + FNEU + FNEG + FPOS + TNEU + FNEG + FPOS + FNEU + TNEG)$$

$$\text{Precision} = [(TPOS \, / \, TPOS + FPOS + FPOS) + (TNEU \, / \, FNEU + TNEU + FNEU) + (TNEG \, / \, FNEG + FNEG + TNEG)] \, / \, 3$$

		Predicted Values		
		Positive	Neutral	Negative
Actual Values	Positive	T_{POS}	F_{NEU}	F_{NEG}
	Neutral	F_{POS}	T_{NEU}	F_{NEG}
	Negative	F_{POS}	F_{NEU}	T_{NEG}

Figure 4. Confusion Matrix

As aforementioned, the evaluation model used in building SA model on Azure ML presents accurate evaluation of the trained model, however, when the model was working on the new datasets, the accuracy was different. In brief, SA model was built on Azure ML using two different algorithms; Logistics Regression and Support Vector Machine. Model was tested using three different datasets (A, B, C). The following subsections summarized the results of the test.

4.1 Testing Azure SA Model with Logistics Regression Algorithm

As shown in Table 1, The sentiment score results of the SA model based on Logistic Regression algorithm for the dataset (A) were 31, 48, 21 for positive, neutral and negative respectively. While for dataset (B), they were 102, 119, 79. Dataset (C) values were 164, 204, 132 respectively. Table 2 shows the confusion matrix for the three datasets (A, B, C) based on the Logistics Regression algorithm with the evaluation metrices.

Table 1. SA Model results - Logistics Regression Algorithm

Logistic Regression	Positive	Neutral	Negative
A	31	48	21
B	102	119	79
C	164	204	132

Table 2. Confusion Matrix based on Logistics Regression Algorithm (Dataset A, B, C)

		Predicted Values								
Actual Values		Positive			Neutral			Negative		
		A	B	C	A	B	C	A	B	C
	Positive	18	47	81	15	41	48	3	11	24
	Neutral	7	39	45	26	70	122	9	25	27
	Negative	6	16	38	7	8	34	9	43	81
Evaluation Metrics										
		A			B			C		
Accuracy		0.53			0.53			0.57		
Precision		0.52			0.53			0.57		

4.2 Testing Azure SA Model with Vector Machine Algorithm

As shown in Table 3, The sentiment score results of the SA model based on Vector Machine Algorithm for the dataset (A) were 37, 40, 23 for positive, neutral and negative respectively. While for dataset (B), they were 87, 127, 86. Dataset (C) values were 196, 175, 129 respectively. Table 4 shows the confusion matrix for the three datasets (A, B, C) based on the Vector Machine Algorithms with the evaluation metrices.

Table 3. SA Model results - Vector Machine Algorithm

Vector Machine	Positive	Neutral	Negative
A	37	40	23
B	87	127	86
C	196	175	129

Table 4. Confusion Matrix based on Vector Machine Algorithm

| | | Predicted Values | | | | | | | | |
| | | Positive | | | Neutral | | | Negative | | |
		A	B	C	A	B	C	A	B	C
Actual Values	Positive	26	51	103	6	27	18	5	18	31
	Neutral	6	21	51	31	87	129	2	29	29
	Negative	5	15	42	3	13	28	16	39	69
Evaluation Metrics										
		A			B			C		
Accuracy		0.73			0.59			0.60		
Precision		0.72			0.56			0.60		

4.3 Sentiment Analysis Models Comparison

In previous sections, the results and the evaluation metrics for each SA model with the three datasets were presented. This section will compare between the two SA models built on Azure ML taking into consideration the Accuracy and Precision attributes.

Figure 4 shows the Accuracy values for the three datasets (A, B, C) based on the Logistic Regression and Support Vector Machine algorithms. Results show that for dataset A, dataset B and dataset, Support Vector Machine algorithm achieved higher Accuracy values at 0.73, 0.59 and 0.60 respectively compared to 0.53, 0.53 and 0.57 for the Logistics Regression algorithm.
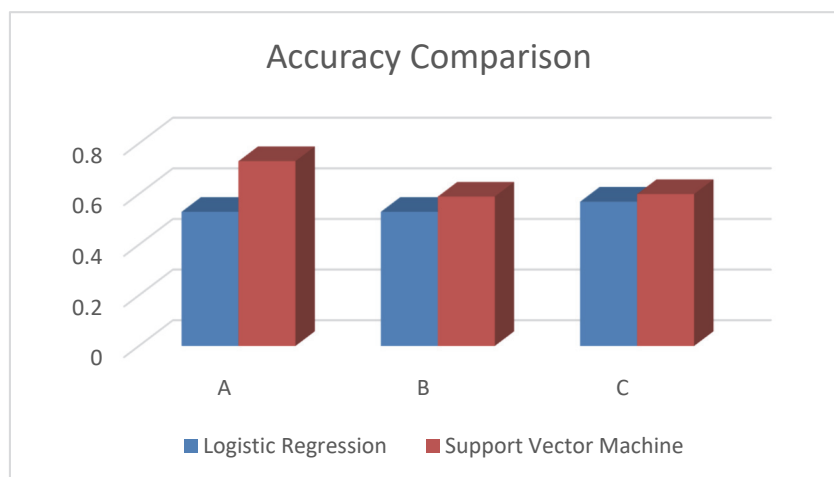

Figure 4. Comparison between two Algorithms based on Accuracy values

On the other hand, Figure 5 shows the Precision values for the three datasets (A, B, C) based on the Logistic Regression and Support Vector Machine algorithms. Results show that for dataset A, dataset B and dataset, Support Vector Machine algorithm achieved higher Accuracy values at 0.72, 0.56 and 0.60 respectively compared to 0.52, 0.53 and 0.57 for the Logistics Regression algorithm.
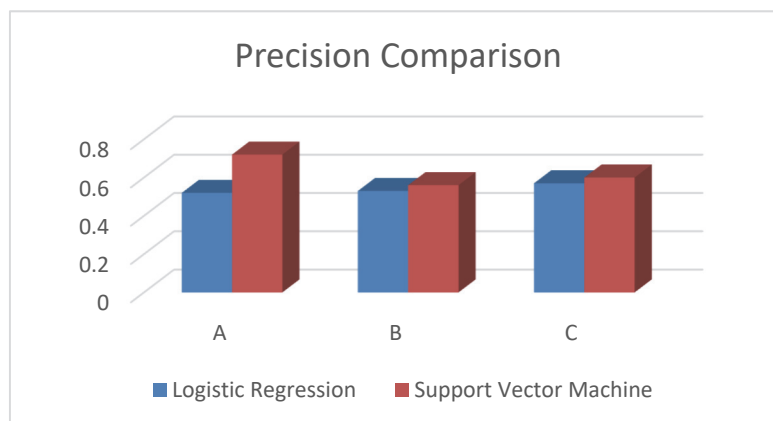
Figure 5. Comparison between two Algorithms based on Precision values

5. Discussion and Conclusion

As demonstrated in this study, sentiment analysis can be built on Microsoft Azure machine language based on two ML algorithms: Logistic Regression and Support Vector Machine (SVM). This is done by performing a series of experiments on three Twitter datasets. Microsoft Azure ML can be used reliably, accurately and securely be used to build SA models. This shows that companies can leverage Microsoft ML to detect customer sentiment and perform topic modelling from several documents. The ability of this service to detect sentiment is achieved using state-of-art learning algorithms that employ scoring attributes and mechanisms when evaluating the text.

Results of this study confirm that prediction models can be implemented on the cloud-based ML platforms. They confirm that SA classification models can be built on cloud-based ML platforms notably Azure ML. This is in line with previous studies that implemented SA systems based on various cloud-based ML platforms (Mulholland et al., 2015; Roychowdhury, 2015; Bornstein et al., 2016).

Similarly, Bihis & Roychowdhury (2015) tested the Generalized Flow performance model built of Azure ML. This model was found to have the ability to perform multi-class and two-class classification. This model was tested using local fundus images dataset, and three Azure datasets: German Credit Card, Wisconsin Breast Cancer, and Telescope. It was revealed that classification accuracy is increased by performing classification based on the Azure ML platform.

In order to identify the quality of SA model that was built using Azure ML, a series of experiments were carried out using different datasets. For assessing the quality of SA models that were built over Azure ML, two evaluation metrics have been used; Accuracy and Precision. Based on these evaluation metrics, this paper stated that building SA models using Support Vector Machine algorithm achieved higher results than Logistics Regression algorithm, though the results were very close for dataset B and C; this research determined that using Support Vector Machine algorithm in building the models attained higher accuracy and Precision values.

As a conclusion, this paper approves that using cloud-based Machine Learning to build Sentiment Analysis models is beneficial, and that is because cloud environment characteristics. Moreover, cloud-based ML platforms are producing reliable models since these platforms are offering users with a set of tools to simplify the process of building the models and to enhance their accuracy.

References

Barbosa, L., & Feng, J. (2010). Robust sentiment detection on twitter from biased and noisy data. *In Proc. of Coling*.

Bihis, M., & Roychowdhury, S. (2015). A generalized flow for multi-class and binary classification tasks: An Azure ML approach. *Big Data (Big Data), 2015 IEEE International Conference*. pp. 1728-1737. Santa Clara: IEEE. https://doi.org/10.1109/BigData.2015.7363944.

Bornstein, A. T., McLoughlin, M. H., Aguilar, J., Wong, W. S., & Solomon, B. D. (2016). Tracking medical genetic literature through machine learning. *Mol. Genet. Metab*, *118,* 255-258. https://doi.org/10.1016/j.ymgme.2016.05.016

Chen, M. Y. (2011). Predicting corporate financial distress based on integration of decision tree classification and

logistic regression. *Expert. Syst. Appl.*, *38,* 11261-11272. https://doi.org/10.1016/j.eswa.2011.02.173

Devika, M., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: a comparative study on different approaches. *Procedia Comput. Sci.*, *87*, 44-49. https://doi.org/10.1016/j.procs.2016.05.124.

Ericson, G., Franks, L., & McKay, P. (2016). What is Azure Machine Learning Studio? Retrieved from Microsoft https://docs.microsoft.com/en-us/azure/machine-learning/machinelearning-what-is-ml-studio

Hasan, D. (2017). A Performance Analysis of Cloud-Based Sentiment Analysis Using Machine Learning, A dissertation submitted in partial fulfilment of the requirements of Sheffield Hallam University.

Kasture, N. R., & Bhilare, P. B. (2016). An Approach for Sentiment Analysis on Social Networking Sites. *In Computing Communication Control and Automation (ICCUBEA) International Conference*, 390-395.

Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, *48,* 354-368.

Liu, B., Zhang, J., Liu, Q., Li, H., Zhang, M., Qiu, R., & Zhao, J. (2015). Data acquisition hot issues and system of microblog mining. *Network and Information Systems for Computers International Conference*, pp 116-119. IEEE. https://doi.org/10.1109/ICNISC.2015.89

Mulholland, E., Kevitt, P.M., Lunney, T., Farren, J., & Wilson, J. (2015). 360-MAM-Affect: Sentiment analysis with the Google prediction API and EmoSenticNet. *Intelligent Technologies for Interactive Entertainment (INTETAIN)*, International Conference, pp. 217-221. Turin: IEEE. https://doi.org/10.4108/icst.intetain.2015.259631

Qaisi, L., & Aljarah, I. (2016). A twitter sentiment analysis for cloud providers: a case study of Azure vs. AWS. *7th International Conference on Computer Science and Information Technology (CSIT)*. ISBN: 978-1-4673-8913-6. Retrieved from https://www.researchgate.net/publication/303811249_A_Twitter_Sentiment_Analysis_for_Cloud_Providers _A_Case_Study_of_Azure_vs_AWS

Qasem, M., Thulasiram, R., & Thulasiram, P. (2015). Twitter sentiment classification using machine learning techniques for stock markets. *Advances in Computing, Communications and Informatics (ICACCI)*. pp. 834-840. India: IEEE. https://doi.org/10.1109/ICACCI.2015.7275714

Rajpurohit, A. (2014). Why Azure ML is the Next Big Thing for Machine Learning? Retrieved from KDnuggets http://www.kdnuggets.com/2014/11/microsoft-azure-machinelearning.html

Roychowdhury, S. (2016). Classification of large-scale Fundus image data sets: a cloud-computing framework. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3256-3259. Lake Buena Vista: IEEE. https://doi.org/10.1109/EMBC.2016.7591423.

Sharef, N. M., Zin, H. M., & Nadali, S. (2016). Overview and future opportunities of Sentiment Analysis approaches for big data. *J. Comput. Sci.*, *12,* 153-168. https://doi.org/10.3844/jcssp.2016.153.168

Thomas, J., McNaught, J., & Ananiadou, S. (2011). Applications of text mining within systematic reviews. *Res. Synth. Methods*, *2*, 1-14.

Topouzelis, K., & Psyllos, A. (2012). Oil spill feature selection and classification using decision tree forest on SAR image data. *ISPRS J. Photogramm Remote. Sens.*, *68,* 135-143. https://doi.org/10.1016/j.isprsjprs.2012.01.005.

Tramèr, F., Zhang, F., Juels, A., Reiter, M. K., & Ristenpart, T. (2016). Stealing machine learning models via prediction APIs. *USENIX Security*. Retrieved from https://arxiv.org/abs/1609.02943.

Vallejos, J., & Mckinnon, S. (2013). Logistic regression and neural network classification of seismic records. *Int. J. Rock Mech. Min. Sci.*, *62*, 86-95. https://doi.org/10.1016/j.ijrmms.2013.04.005

Wu, Z., Zhang, H., & Liu, J. (2014). A fuzzy support vector machine algorithm for classification based on a novel PIM fuzzy clustering method. *Neurocomputing*, *125,* 119-124. https://doi.org/10.1016/j.neucom.2012.07.049

Copyrights