sjst.scst.edu.ly

Surman Journal of Science and Technology

ISSN: Online (2790-5721) - Print (2790-5713)

مجلة صرمان للعلوم والتقنية

Vol. (4), No.(1), May_2022

Pages: 001 ~ 014

Using Artificial Neural Network as an Approach to Analyze People Sentiment Level Based on Social Media Data

Khalid J. Bisher[1], Abubaker Kashada[2], Riyadh A. Alsayih[3], Mohamed Alforgani[4], Eenas A. Suwayd[5]

[1, 2, 3] Surman College of Science and Technology

[4] Sok Jomma Institute for Technical Science, Tripoli, Libya

[5] Higher Institute of Science and Technology, Sabratha, Libya

[1]*kalid_bisher@scst.edu.ly*

[2]*kashada@scst.edu.ly*

[3]*riyad_alsaeh@ scst.edu.ly*

[4]*mabdunnabi1970@gmail.com*

[5]*eenas_suwayd@shist.edu.ly*

Abstract

The internet has become an essential online communication tool for many people today. For a variety of reasons, many researches have been done lately in the field of Artificial Neural Network (ANN) considering people sentiments based on social media data. This analytical study conducted to analyze happiness of Libyan people based on twitter data sets.

Matlab used in this study to code around 1,000 status/comments data. Processing consists of five processes, namely cleansing, Tokenization, case folding, removal stop word, and stemming. The study represents Artificial Neural Network model for the mining of Twitter opinions using an Artificial Neural Network model approach for the abstracting and visualization scheme of Twitter feeds and a classification and prediction approach. This study presented a contribution in the form of proposing a new visualization model for Twitter mood prediction based on the ANN approach.

Keywords: *Artificial Neural Network, People Sentiments, Social Media Data, Twitter Data analysis*

1. Introduction and Background

As a result of information and communications technology developing and the very rapid spread and growth, internet have become part of everyday life and become an essential online communication tool [1][2]. People use social media networks to find friends, sharing moments, ideas, and even exist to run online businesses. There are many social networking sites on the internet, such as Twitter, Facebook, Linkedin, Instagram, and Google + [3][4]. Twitter is a social media that is widely used nowadays. It provides several pieces of information about someone's sentiment. Too many people are using such media to express opinions as well as emotions and expressing their emotions like happiness, angry, sadness, etc... The use of Twitter was evidenced by the existence of many studies have used Twitter as material to analyze the sentiments of someone's opinion.

Several classifications studies conducted using Artificial Neural Network ANN such as a study conducted by Leavline [5] to classify the image of orchids to distinguish plant variations. Also, the ANN method was used by Wechmongkhonkon, et al. [6] to classify water levels to manage water quality. A study conducted by Alaloul, et al. [7] used ANN simulation to estimate hospital building cost from the building contract. During the planning stage with the best ANN structure, 7-9-1 (7 input variables, one hidden layer with nine neurons, and one output) results from an average cost estimation of 96.51% achieved. An adaptation of three multi-layer feed-forward networks with Back-Propagation and Elman-Propagation algorithms to test the time, quality and cost, and to fulfil, the objective of validate, training as performance evaluation indicators. As a result, the training process is a constant flow unit until it works its way to the

predefined error or up to 1000 epochs. The outcome is presented in a Mean Square Error (MSE) confirmed the accuracy of the networks with an average value of 0.0231. A study conducted by Nwankpa, et al. [8] that was designed to implement the KNN (K-Nearest Neighbor) algorithm for sentiment analysis twitter about Jakarta Governor Election 2017. The data was collected from Indonesian Tweets presented in 2000 tweet in particular; the tool was used is Python package called Twitters craper. KNN is the method used for sentiment analysis combined with Cosine similarity measure and TF-IDF term weighting. The results indicated that the level of highest accuracy is 67, 2% when k=5, the highest precision is 56,94% with k=5, and the highest recall 78,24% with k=15.

The above-mentioned approach showed that the classification method using the ANN is considered to give quite good results [9]. This study adopted Artificial Neural Network as a method to classify document data. ANN is widely used in computer vision problems for pattern recognition and image classification. this study used a text mining analysis of a person's sentiment from a tweet. Twitter use is increasing by the year and become more common as the time progresses [10]. Figure (1) illustrates the growth rate forecast of Twitter users.

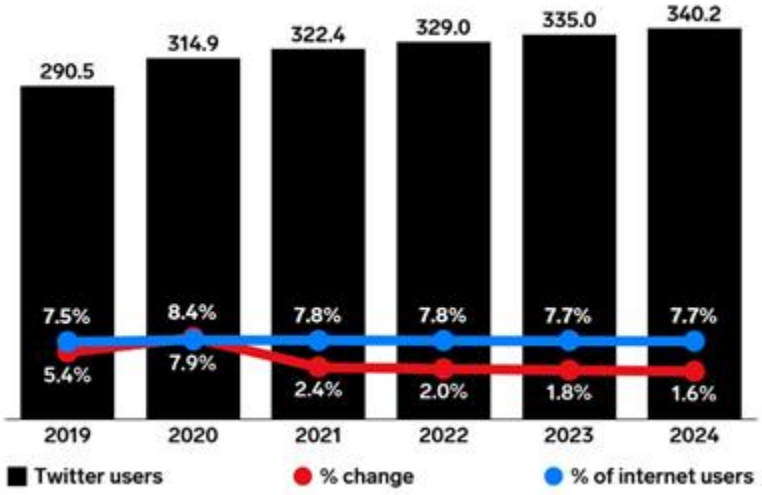

Figure.1 Twitter users growth worldwide, 2019-2024
Image Source: https://www.emarketer.com

1.1 Artificial Neural Network

Artificial neural networks are first revealed by Walter Pitts and Warren MacCulloh in 1943 [11]. Both designed a simple calculation model that unites biological science related to neural networks in humans with the science of logic in solving a problem. Artificial neural networks that are made are computational-based artificial intelligence analogous to how the human neural network system works, as in Figure 2.1 [12].

Surman Journal of Science and Technology
ISSN: Online (2790-5721) - Print (2790-5713)
مجلة صرمان للعلوم والتقنية
Vol. (4), No.(1), May_2022
Pages: 001 ~ 014
sjst.scst.edu.ly

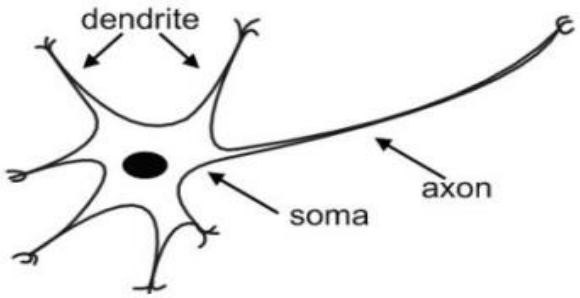

Figure 2. Human Neural Network
Image Source: Leavline, E. [5]

Artificial neural networks that are analogous to human neural networks can be present by the mathematical model shown in Equation (1):

$$a = \sum_{j=1} wj \ wk \ \dots\dots\dots\dots\dots\dots\dots (1)$$

Where
xj = binary input to j
wk = sympathetic weight
The logic that builds Eq. (1) is shown in Figure 3.

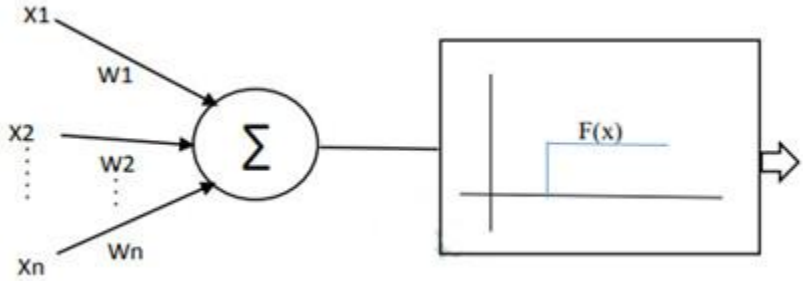

Figure 3. ANN function
Image Source: Leavline, E. [5]

$wn =$ Synaptic weight is a real number value of related synapses [3a]. Synapses can inhibit negative signals or activate positive signals. Besides, synapses also have different intensities. The a signal is the input of the activation function F (x) that gives neuron output [5]. In 1958 Rosen Boltt developed a network model to optimize the results of the iteration [13].

1.2 ANN classification

Basically, ANN classification is undertaken to achieve the purpose of pattern recognition trained to identify problems that are given in phase 2 phase, namely training and testing. During training process that occurs, the network is trained to classify data based on models that have been built during training to produce 0% errors at the validation stage [14]. Artificial neural

networks have several different algorithms that are used in completing forecasting or prediction in a study, including Single network and Multi-Layer Network [15]. This study attempts to conduct an analytical study of data based on social media explicitly using Twitter data to analyze happiness of Libyan people based on Twitter data using artificial neural network.

2. Research Problem & Questions

Text in comments or status on Twitter can indicate that someone is happy or unhappy. Thus, the study problem was, how to identify Libyan happiness level based on Twitter text using an artificial neural network. Based on the argument above, the research questions could be as follows: How to analyze happiness of Libyans using the artificial neural network based on Twitter data?

3. Research Conceptual Model

The sentiment analysis architecture shows in figure 4:

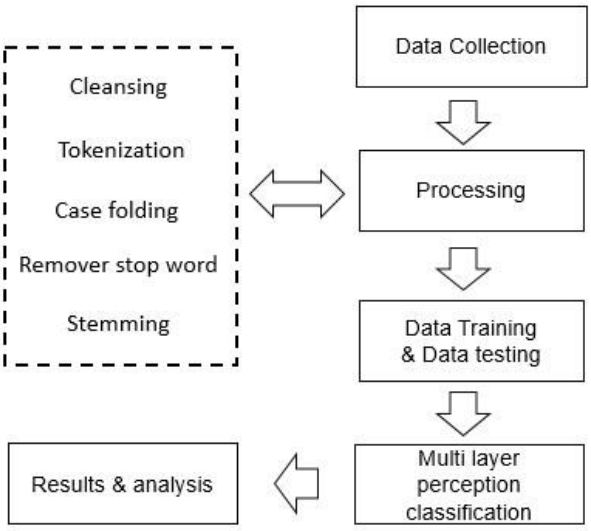

Figure 4. Research Conceptual Model

4. Data Collection

The first method captures Twitter data directly from within the Twitter service via a Twitter archive request. The interactive interface is the data that can be downloaded using the HTML index page, JSON data set, and a CSV file that contains all tweets. A significant data is provided from the archival function these sufficient data forms the primary ground for the process of internal analysis, the objective and the primary purpose of this work is to examine the individual, personally controlled Twitter timeline. Data collected in this study using social media especially for Libyan people. Matlab used in this study to code around 1,000 status/comments data from a different accounts of Libyan people.

5. Data processing/analysis

processing consists of five processes, namely cleansing, Tokenization, case folding, removal stop word, and stemming as shown in figure 5.

sjst.scst.edu.ly
Surman Journal of Science and Technology
ISSN: Online (2790-5721) - Print (2790-5713)
مجلة صرمان للعلوم والتقنية
Vol. (4), No.(1), May_2022
Pages: 001 ~ 014

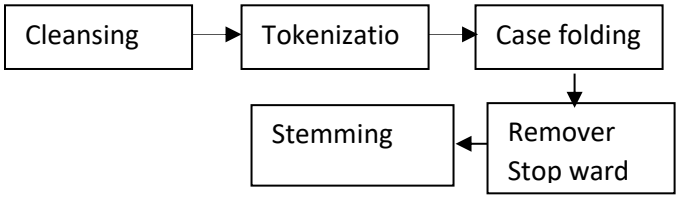

Figure 5. Data processing

5.1 Validation, Training, and Test Datasets

Matlab software used to separate dataset into three parts as the following: training, testing, and validation data. Where, training Dataset is the sample of data used to fit the model, testing dataset is the sample of data used to provide an unbiased evaluation of a final model fit on the training dataset, and validation Dataset is the sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters.

5.2 Pre-processing

The table below shows the data preparation:

Table 1. Example of Data Preparation

ID Data	Twitter Text	Input Layer								Sentiment (happiness)
		Happy	sad	kind	bad	peace	angry	…	f_m	
1	Happiness is warm puppy	1	0	0	0	0	0			1
2	Happiness start today	1	0	0	0	0	0			1
3	I feel very sad	0	1	0	0	0	0			0
4	Kindness and happiness	1	0	1	0	0	0			1
5	Happiness inspiration	1	0	0	0	0	0			1
6	Bad day	0	0	0	1	0	0			0
7	Make me angry	0	0	0	0	0	1			0
8	Stay Home, happiness	1	0	0	0	0	0			1
9	Unhappiness and rain just drift by and are gone	0	0	0	0	0	0			0
…										
1000	My friend is very kind and make me in peace	0	0	1	0	1	0			1

n = number data (1000 data) | m = number of selected feature | sentiment = 1 (happy) | sentiment = 0 (unhappy) | Word 1 = token (ex. Pleasant, happy, optimize, bad, angry, etc.)

5.3 Mapping ANN Architectural for Happiness Index

The Figure 6. below illustrates a typical architecture of the feed-forward ANN happiness model structure:

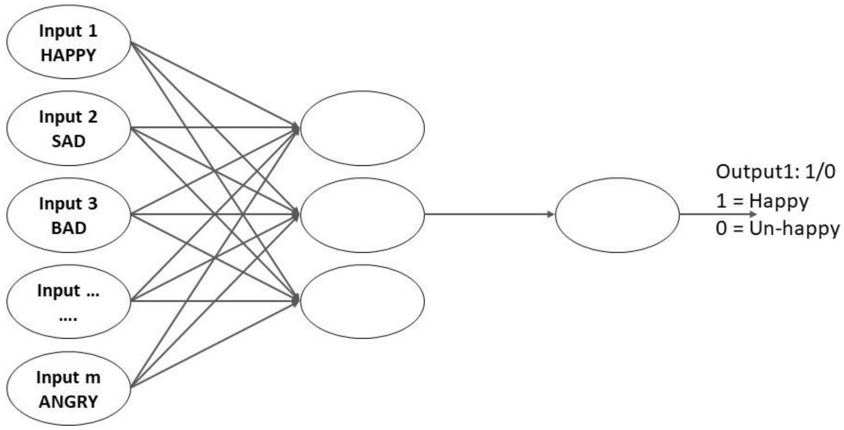

Figure 6. Input and Output by ANN

Happiness input 1,2,3 is the cleaning step process. Happiness input layer is the tokenization process. Happiness hidden layer is the case folding and omission step process. Happiness output layer is the multi-layer alogaritms.

5.4 Analysis Flowchart

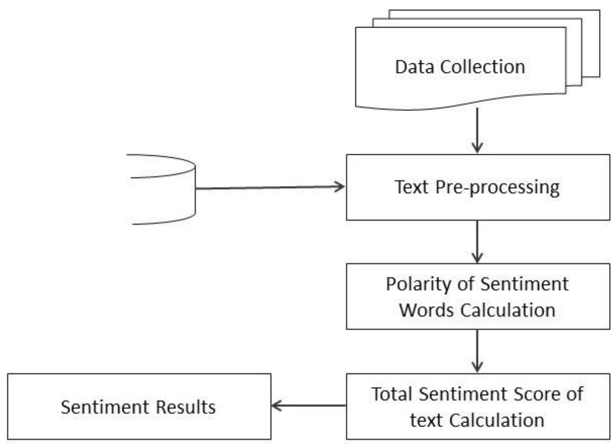

Figure 7. Sentiment Analysis Flowchart

The sampling methods is a method of data collection. Conducted by using a text processing method, to calculate the polarity of sentiment words, the formula or Equation that is used:

$$MSE = \frac{\sum_{j=1}^{n}(oi-ti)^2}{n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

in which, (Oi) is the targeted output for the training data or cross-validation data (i), (Ti) represents the network output for the training data or cross-validation data (i), and (n) is the number of data in the training data set or the cross-validation data set [16]. The relative importance of an input parameter i is determined using the following equation:

sjst.scst.edu.ly
Surman **J**ournal of **S**cience and **T**echnology
ISSN: Online (2790-5721) - Print (2790-5713)
مجلة صرمان للعلوم والتقنية
Vol. (**4**), No.(**1**), May_2022
Pages: 001 ~ 014

$$\operatorname{Im} p(i) = \sum_{i=1}^{n} (CW_{ih(x)}\, CW_{ho(x)} \quad \ldots\ldots\ldots\ldots \quad (3)$$

where Imp(i) is the relative importance of parameter (i), (n) is the total number of hidden neurons, (x) is the index number of the hidden neuron, CWih(x) is the connectivity weight between input parameter (i) and the hidden neuron (x) and CWho(x) is the connectivity weight between the hidden neuron (x) and the output neuron. In this study, Eq. (4) was used to determine the importance of parameters in our ANN [17].

The flowchart of sentiment analysis contains the product review, sentiment identification, feature selection, sentiment classification and polarity. The sentiment analysis is divided into the machine learning approach and lexicon-based approach. The machine approaches further divided into supervised learning and unsupervised learning. Meanwhile, the lexicon-based approach contains both a dictionary-based approach and a corpus-based approach. The sentiment analysis performs and start from the data collection and text preprocessing, after that calculate of sentiment words, and make sentiment results [18].

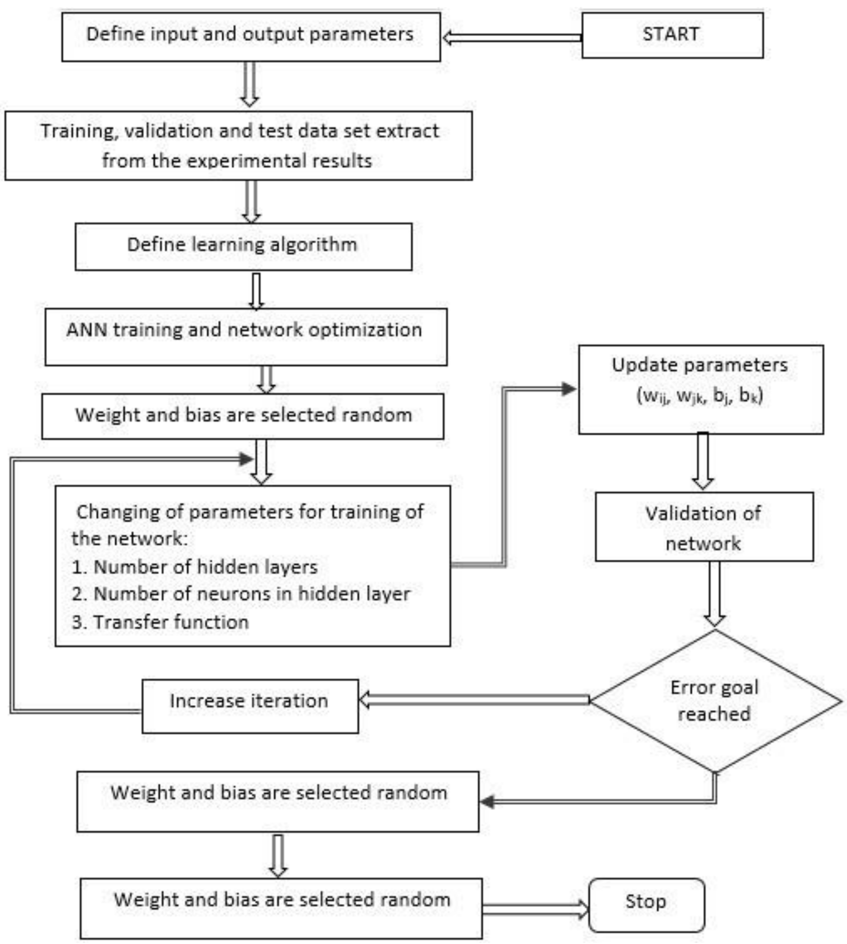

Figure 8. Research Flowchart

sjst.scst.edu.ly
Surman Journal of Science and Technology
ISSN: Online (2790-5721) - Print (2790-5713)
مجلة صرمان للعلوم والتقنية
Vol. (4), No.(1), May_2022
Pages: 001 ~ 014

6. Results and Discussion

A sample of 1000 responses from Twitter among Libyan that express their feeling such as happy and unhappy will be analysis, that consists of text cleaning from data collected, labelling, dataset and result from analysis.

The number of input in an artificial neural network (ANN) 323 and the number of hidden layers used was 2.

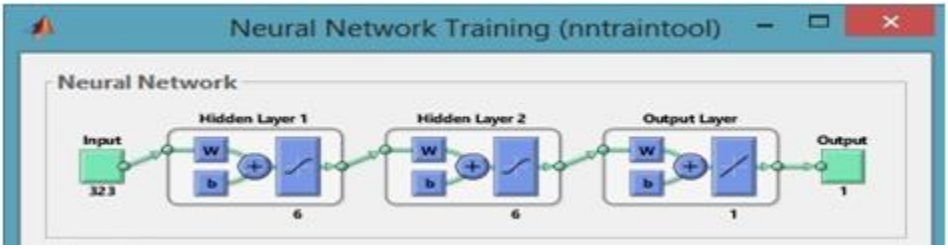

Figure 9. Input and hidden layers

Table 2. Parameter Multilayer Perceptron

No	parameter	information
1	Input layers	323
2	hidden layers 1	6 node
3	hidden layers 2	6 node
4	Output layers	1

This is a sample of the data collected and analyzed and the tables [2] show examples of cleansing, data labelling, tokenization, case folding and stemming analysis.

Table 3. Data Clean Message

No	Message	Label	class	Clean Message
1	Life is an awful, ugly place to not have a best friend.	Unhappy	0	Life is an awful ugly place to not have a best friend
2	he is smiling at you	Happy	1	he is smiling at you
3	I tired to be with you, seriously	Unhappy	0	I tired to be wiht you seriously

Table 4. Data Labelling

No	Message	Label	Class
1	Life is an awful, ugly place to not have a best friend	unhappy	0
2	he is smiling at you	happy	1
3	I tired to be wiht you seriously	unhappy	0

Table 5. Data Tokenization

Clean Message	Life is an awful ugly place to not have a best friend
	he is smiling at you
	I tried to be with you seriously
Tokenization	Life|is|an|awful|ugly|place|to|not|have|a|best|friend
	he|is|smiling|at|you
	I|tired|to|be|with|you|seriously

Table 6. Data Case Folding

Clean Message	Life is an awful, ugly place to not have a best friend
	he is smiling at you
	I tried to be with you, seriously
Tokenization	Life|is|an|awful|ugly|place|to|not|have|a|best|friend
	he|is|smiling|at|you
	I|tired|to|be|with|you|seriously
Case Folding	life|is|an|awful|ugly|place|to|not|have|a|best|friend
	he|is|smiling|at|you
	i|tired|to|be|with|you|seriously

Table 7. Data Source Steaming

Case Folding	life|is|an|awful,|ugly|place|to|not|have|a|best|friend
	he|is|smiling|at|you
	i|tired|to|be|with|you,|seriously
Stop Word Removal	life|awful|ugly|place||not|have|a|st|friend
	he|smiling|at|you
	i|tired |with|you|seriously
Stemming	life|ful|ugly|place||not|have|a|st|friend
	he|smile|at|you
	i|tired |with|you|seriously

6.1 Source Feature Selection and data set

Table 8. Data Source Feature Selection

No	Clean Message	Tokenization	Case Folding	Stop Word Removal	Stemming	Selected Feature
4	Two wrongs don't make a right, but they make a good excuse	Two|wrongs|don't|make|a|right|but|they|make|a|good|excuse	two|wrongs|don't|make|a|right|but|they|make|a|good|excuse	two|wrongs|don't|make|a|right|but|they|make|a|good|excuse	two|wrongs|don't|make|a|right|but|they|make|a|good|excuse	good
5	Life is an awful, ugly place to not have a best friend.	Life|is|an|awful|ugly|place|to|not|have|a|best|friend|	life|is|an|awful|ugly|place|to|not|have|a|best|friend|	life|is|an|awful|ugly|place|to|not|have|a|best|friend|	life|is|an|awful|ugly|place|to|not|have|a|best|friend|	ugly
6	For every minute you are angry, you lose sixty seconds of happiness.	For|every|minute|you|are|angry|you|lose|sixty|seconds|of|happiness	for|every|minute|you|are|angry|you|lose|sixty|seconds|of|happiness|	for|every|minute|you|are|angry|you|lose|sixty|seconds|of|happiness|	for|every|minute|you|are|angry|you|lose|sixty|seconds|of|happiness|	angry

Table 9. Data Set

feature 1	feature 2	feature 3	feature 4	feature 5	feature 6	feature 7	feature …	feature 323
hungry	lie	happy	good	ugly	angry	soul		just
1	0	0	0	0	0	0		0
0	1	0	0	0	0	0		0
0	0	1	0	0	0	0		0
0	0	0	1	0	0	0		0
0	0	0	0	1	0	0		0
0	0	0	0	0	1	0		0
0	0	0	0	0	0	1		0
0	0	0	0	0	0	0		0

6.3 Classification using the Perceptron Multilayer Algorithm

Multi-layer perceptron algorithm (MLP) was used in classification by using the obtained tweet's term. Every word with the highest weight was used as a value for IDF value.

sjst.scst.edu.ly
Surman Journal of Science and Technology
ISSN: Online (2790-5721) - Print (2790-5713)
مجلة صرمان للعلوم والتقنية
Vol. (4), No.(1), May_2022
Pages: 001 ~ 014

Figure 10. Neural Network Training

Neural Network training showed revealed 13 iterations with six times validation with error goal (MSE) 0.00100. The Matlab output shows that the higher happy mood will affect the complimentary sentiment analysis of ANN in Libyan Twitter.

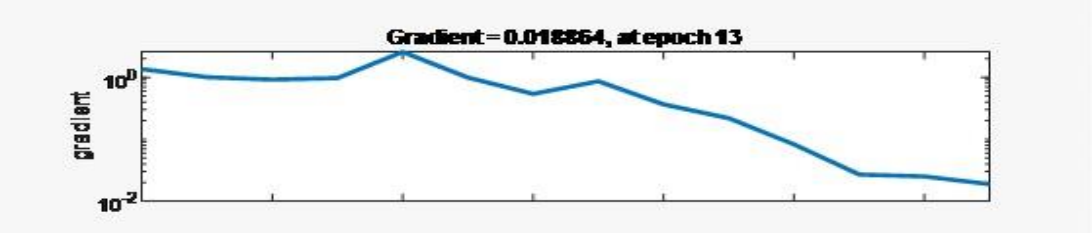

Figure 11. Gradient

Figure 12. Shows result of gradient, is 0.018864 at epoch 13 and tend to be decreased.

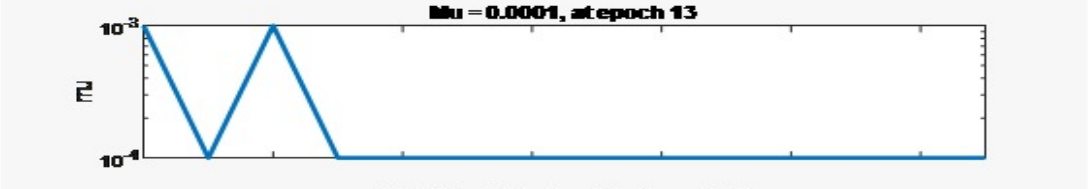

Figure 12. Result of Mu

Figure 13. Shows the result of Mu, is 0.0001 at epoch 13 and tend to be decreased.

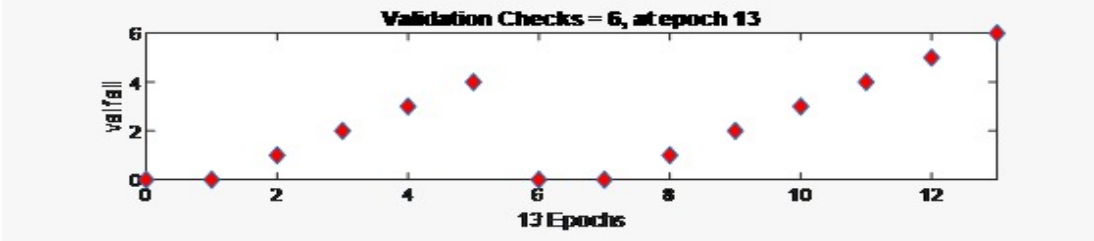

Figure 13. Validation checks

Figure 14. Shows that the validation checks is 6, at epoch 13, and tend to be decreased.

Fuzzy Mamdani and Fuzzy Sugeno were used in order to validate the results of ANN method, as shown in Figure 14.

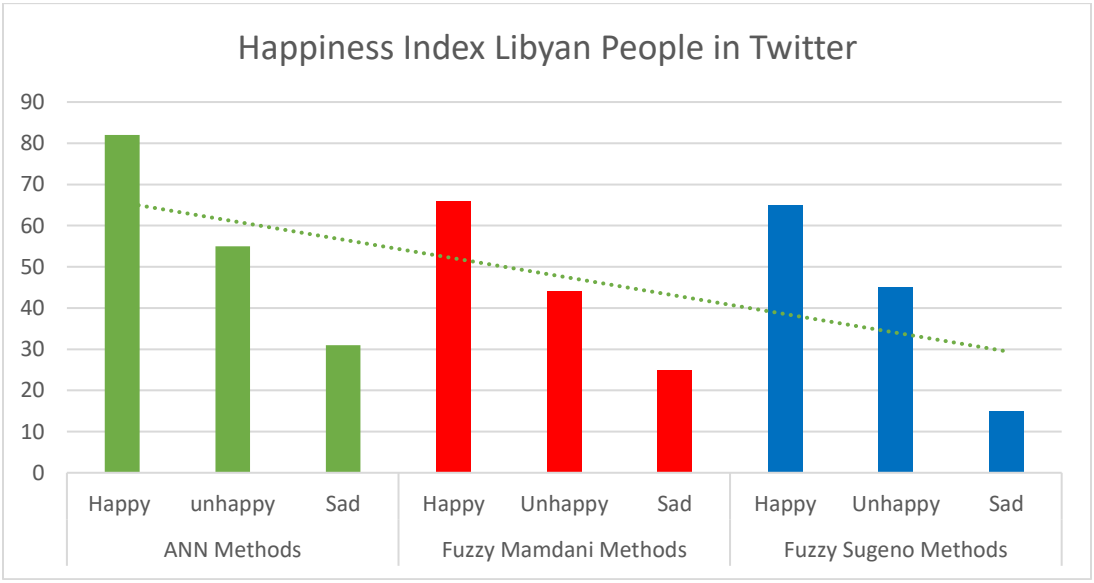

Figure 14. Comparative analysis of Happiness Index between ANN, Mamdani and Sugeno Methods

The result of validation verified that ANN method has perform with a good results compared to Fuzzy Mamdani and Fuzzy Sugeno. The percentages of the performance index of ANN has 25% more higher compared with other methods.

7. Conclusion

The research study was an analytical study of data based on social media explicitly using Twitter data. Life became to rely on social media networks in a significant way, amongst all social media available, Twitter considers as the most popular used media regardless of the joint opinion towards social media For a variety of reasons, many researches have been done lately in the field of Artificial Neural Network (ANN) considering people sentiments based on social media data, for social media analysis, Twitter is a good starting point, the reason for that is users tend to share their opinions to general and publicly. Artifical Neural Network were successfully carried out on Twitter sentiment level analysis. Negative and positive sentiment values provided as system output. The research contributes by presenting visualisation model for Twitter mood prediction and represents an Artificial Neural Network model for Twitter opinions mining using a classification and prediction approach.

6. References

[1]. Kashada, Abubaker, and Wesam Mohamed AllaEddinGhaydi. "The impact of perceived usefulness & perceived ease of use on the successful adoption of information systems in developing countries." IOSR Journal of Computer Engineering (IOSR-JCE) 22.1 (2020): 45-48.

[2]. Kashada, Abubaker, Hongguang Li, and Osama Koshadah. "Analysis Approach to Identify Factors Influencing Digital Learning Technology Adoption and Utilization in Developing Countries." International Journal of Emerging Technologies in Learning (iJET) 13.02 (2018): 48-59. https://doi.org/10.3991/ijet.v13i02.7399

[3]. Kashada, Abubaker, Hongguang Li, and Osama Kashadah. "The impact of user awareness on successful adoption of decision support System DSS in Developing Countries: THE Context of Libyan Higher Education Ministry." American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS) 16, no. 1 (2016): 334-345.

[4]. Kashada, Abubaker, and Wesam Mohamed AllaEddinGhaydi. "The impact of perceived usefulness & perceived ease of use on the successful adoption of information systems in developing countries." IOSR Journal of Computer Engineering (IOSR-JCE) 22.1 (2020): 45-48.

[5]. Leavline, E. Jebamalar. "Classification Problem Using MATLAB." nternational Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST) 1.6 (2015): 22-25.

[6]. Wechmongkhonkon, S., N. Poomtong, and S. Areerachakul. "Application of artificial neural network to classification surface water quality." World Academy of Science, Engineering and Technology 6.9 (2012): 574-578.

[7]. Alaloul, Wesam Salah, et al. "An Artificial neural networks (ANN) model for evaluating construction project performance based on coordination factors." Cogent Engineering 5.1 (2018): 1507657.

[8]. Nwankpa, Chigozie, et al. "Activation functions: Comparison of trends in practice and research for deep learning." arXiv preprint arXiv:1811.03378 (2018).

[9]. Lakshmi Praveena, T., and N. V. Muthu Lakshmi. "Predicting Sentiment and Emotion on ASD Using Machine Learning and Deep Neural Networks Based on Twitter Data." Intelligent Sustainable Systems. Springer, Singapore, 2022. 473-487.

[10]. Chandrasekaran, Ganesh, et al. "Visual Sentiment Analysis Using Deep Learning Models with Social Media Data." Applied Sciences 12.3 (2022): 1030.

[11]. Merayo, Noemí, et al. "PID controller based on a self-adaptive neural network to ensure QoS bandwidth requirements in passive optical networks." Journal of Optical Communications and Networking 9.5 (2017): 433-445.

[12]. Chaabouni, Rahma, et al. "Communicating artificial neural networks develop efficient color-naming systems." Proceedings of the National Academy of Sciences 118.12 (2021).

[13]. Santos, Iria, et al. "Artificial neural networks and deep learning in the visual arts: A review." Neural Computing and Applications 33.1 (2021): 121-157.

[14]. Li, Quanzhi, et al. "Tweet sentiment analysis by incorporating sentiment-specific word embedding and weighted text features." 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI). IEEE, 2016.

[15]. Abdesselam, Abdelhamid, and Rahmat Choo Abdullah. "Pepper berries grading using artificial neural networks." 2000 TENCON Proceedings. Intelligent Systems and Technologies for the New Millennium (Cat. No. 00CH37119). Vol. 2. IEEE, 2000.

[16]. Roudatul, Zannah. Analisis sentimen pada media sosial Twitter untuk klasifikasi opini Islam Radikal menggunakan Jaringan Saraf Tiruan. Diss. Universitas Islam Negeri Sunan Ampel Surabaya, 2019.

[17]. Li, Zhenjun. "A Data Classification Algorithm of Internet of Things Based on Neural Network." International Journal of Online Engineering 13.9 (2017).

[18]. Rachmat, Antonius, and Yuan Lukito. "Sentipol: Dataset Sentimen Komentar Pada Kampanye Pemilu Presiden Indonesia 2014 Dari Facebook Page." Konferensi Nasional Teknologi Informasi dan Komunikasi 2017 (2016): 218-228.