

Rocchio Nearest Centroid and Normalized Neural Network based Lead Generation in Social Media Marketing

K.S. Narayanan¹; Dr.S. Suganya²

¹Associate Professor, Rathnavel Subramaniam College of Arts and Science, Coimbatore, India.

¹narayanan@rvsgroup.com

²Associate Professor, Rathnavel Subramaniam College of Arts and Science, Coimbatore, India.

²suganya_cs@rvsgroup.com

Abstract

The employment of the internet and social media has transposed consumer behavior and the methods in which business organizations carry over their business. Social and digital marketing recommends noteworthy changes to business establishments via cost curtailment, enhanced brand understanding and surged sales. Despite enormous amount of potentialities, noteworthy disputes subsist from gloomy digital oral message in addition to trespassing and troublesome online presence of brand. Deep learning (DL) has fascinated escalated awareness owing to its notable processing power in tasks, to name a few being, speech, image, or text processing. Due to its aggressive evolution and extensive accessibility of digital social media (SM), examining these data utilizing conventional materials and methods is substantial or even complex. Also with the large growth in the volume of data, the multifariousness in data heterogeneity, are the most distinguished reasons, why and how the SM data mounted. In this paper we study the impact of tweets on distance learning to understand people's opinions and to discover facts. However, adding redundant features minimizes the generalization capability of the model and may also minimize the overall accuracy of a classifier. We introduce Rocchio Nearest Centroid Laplacian Feature Selection model that combines Rocchio Nearest Centroid and Laplace function for selecting relevant features or tweets. Next an Arbitrary Normalized Attention-based Recurrent Neural Network Lead Generation algorithm is designed aggregating characterizations from preceding and succeeding tweets while generating lead via digital marketing tweet funnel. We validate and evaluate our method using data from distance learning dataset. Experiments and comparisons on distance learning data show that, compared to existing SMM methods, considering generalization capability and digital marketing tweet funnel results in improvements in processing time, lead generation accuracy and precision to a significant extent.

Key-words: Deep Learning, Social Media, Rocchio, Nearest Centroid Laplacian, Arbitrary Normalized, Attention-based Recurrent Neural Network.

1. Introduction

As far as digital advertising is concerned, user interest and behavior modeling are considered to be the most interpretative stride. On one hand, user interests have straight influence on the response and actions to Displayed Advertisement (Ad). On the other hand, user interests can supplementary assist in evaluating the potentiality of an Ad viewer converting into a potential customer. So far, prevailing methods for digital advertising specifically think about constituting users as an invariable feature set and train learning classifiers for efficient prediction. However, those methods do not take into consideration temporal discrepancy and alternates in user behaviors, and completely depend on provided features for learning.

Two Deep Learning-based frameworks, called, LSTM-cp and LSTM-ip, considering user click prediction (cp) and user interest modeling for user click prediction and user interest prediction was proposed in [1]. The objective of the work remained in precisely predicting the potentiality of a user making a robust clicking on an Ad and the potentiality of user clicking on a definite type of advertisement campaign.

To attain the objective, information pertaining to the page was collected from the users as a temporal sequence and utilize Long Short Term Memory (LSTM) network to learn features that constitutes user interests as latent features, therefore improving precision, recall and prediction accuracy. However, adding redundant features minimizes the *generalization capability* of the model and may also minimize the overall accuracy of a classifier. To address this issue, a Rocchio Nearest Centroid Laplacian Feature Selection model is first proposed that with the Rocchio classifier addresses the *generalization capability* by means of Laplacian score, therefore contributing to both accuracy and processing time to a greater extent.

In [2], ML integrated Social Media Marketing (ML-SMM) was proposed with the objective of obtaining ceaseless revenue and acquiring several customers using data mining techniques. Three distinct steps were utilized in the design of ML-SMM method. They were mining the textual data, integration of machine learning with social media marketing, and finally, the performance analysis of ML-SMM with the aid of WEKA. With this the true positive, false positive, precision and recall involved in social media data analytics were found to be improved.

However, research works conducted in SMM vigorously recommend the existence of a media marketing funnel, i.e., a customer may go through numerous stages before finalizing a purchase (conversion). If this media marketing funnel results are not properly analyzed precision is said to be compromised. To address this issue, in this work, Rocchio Nearest Centroid and Normalized Neural

Network (RNC-NNN) lead generation model is designed that with the aid of arbitrary normalized attention function create a two-way direction between the preceding and succeeding via neural network. With this, the precision involved is improved.

Our contributions summarized as follows:

- We propose a model to generate lead and rank the polarity score on SMM and applied on distance learning dataset to understand people's opinions and to discover facts by analyzing the positive and negative tweets while selecting course in universities.
- This study focuses on a given time point along with acquiring the lead and makes an effort to select the relevant feature which may get great attention of the users or student's community based on the tweet polarities.
- Proposed method makes prediction of already popular tweets about distance learning with minimum processing time and maximum accuracy by means of Rocchio Nearest Centroid Laplacian Feature Selection.
- Design Arbitrary Normalized Attention-based Recurrent Neural Network Lead Generation model for generating the lead with maximum precision.
- We bestow significant empirical investigations on the chosen dataset to support obtained results. The said datasets contain the latest tweets provided by the student community about distance learning. We then model these networks to generate the lead with maximum precision.

The remainder of this paper is organized as follows. Related works is described in section 2, followed by section 3 explaining our overall method. We then cover the problem definition, followed by feature selection and lead generation model. The experimental results on both feature selection and lead generation are covered in section 4. Section 5 discusses in detail regarding the processing time, lead generation accuracy and precision concerning lead generation method. Section 6 covers the conclusion.

2. Related Works

In the near future, artificial intelligence (AI) is said to evidently transform both marketing techniques and customer behaviors. Designing from not only enduring research but also sizeable relationships with implementation, the authors presented a multidimensional technique apprehending the influence of AI necessitating intelligence intensities, task categories and the embedding of AI in a robot for future marketing [3].

A Technology Acceptance Model (TAM) was proposed in [4] to analyze the comprehensive cognizance of students' purpose to utilize digital mechanisms in an amalgamated learning context of higher education. This mechanism was also found to be reliable and invariant across degree levels.

In [5], an elaborate and timely contribution was presented to both research community and the practitioners in the mode of ultimatums and chances where the pinnacle was also made on the disadvantages in assisting within the purview of both digital and social marketing. An advanced architecture and workflow for analyzing comprehensive data obtained from social media streams was proposed and investigated in [6]. A predictive machine learning approach based on bayes theorem was used for classifying the flood and no flood messages.

One of the promising research areas' to evaluate user's ideology from their past pursuits is intention mining. These past pursuits are stored in the form of logs via search engines. The search engines in turn ease the sellers and producers to bestow their products to the end user in an encouraging manner. A systematic literature review for social media intention mining was proposed in [7].

Though online marketing has started spreading faster than the manual marketing misinformation are found in large number that in turn hinders the accuracy of content being retrieved. A survey of misinformation detection system using deep learning was proposed in [8]. In [9], a Bidirectional Encoder Representations from Transformers–based (BERT) deep learning approach called (FakeBERT) was proposed by integrating distinct parallel blocks of single-layer deep Convolutional Neural Network (CNN) possessing numerous kernel sizes and filters, therefore contributing to classification accuracy.

Despite key role played in associating people globally, the digital social media provides an enormous type of knowledge extracting tasks. Extracting the most trivial information and gaining insight from this trivial information is not even yet an insignificant issue to address. Machine learning methods, chaperone with the advancement in prevailing computing power, take part a predominant part to hold disguised information in this data.

A novel Long Short Term Memory (LSTM) and Convolutional Neural Networks (CNN)–grid search-based deep neural network model for sentiment analysis was proposed in [10]. A taxonomy oriented summary on representing the social media analytics oriented issues was proposed in [11]. Poll opinion of public regarding demonetization via Aspect Based Sentiment Analysis (ABSA) was presented in [12]. Here, an integrated meta-heuristic algorithm combining fire fly and multi verse was proposed for efficient classification of positive and negative sentiments. With this classified results, the competent performance for demonetization tweets were analyzed.

Yet another emotion analysis using deep learning for sentence classification was proposed in [13]. In [14], a three stage strategic framework for examining human emotions towards social media marketing was proposed, therefore to arrive at decisions and feelings. With the high-speed evolution of web content from social media, like, online opinion mining has started gaining awareness from government institutions, industry, academic sectors and so on. In recent few years, analysis of sentiment has not only made an appearance underneath knowledge integration in the big data epoch, but has also become an accepted research area in the domain of artificial intelligence and machine learning.

A study was conducted in [15] utilizing the Military life PTT board of Taiwan's largest online forum. The purpose remained efficiently designing a sentiment analysis model for social media towards enhancing sentiment classification via deep learning models with numerous parameter combinations, therefore improving the accuracy of sentiment polarity judgement. Sparse recurrent neural networks were employed in [16] to further quicken the training process with adaptive connectivity.

Consumer behavior analysis based on neural network was proposed in [17] towards measuring and understanding the consumer behavior in international marketing. Detection of hate speech in social media was analyzed by means of deep natural language processing model to efficiently learn various features [18] integrating convolutional and recurrent layers. With this the results obtained show significant increase in the classification score. The effectiveness of content marketing and findings from managerial angle was analyzed in [19].

Motivated by the above findings, social media marketing using feature engineering, called, Rocchio Nearest Centroid and Normalized Neural Network (RNC-NNN) is proposed in our work that in turn lay foundation for institutions to digitally promote higher education to reach the potential customers via tweets.

3. Rocchio Nearest Centroid and Normalized Neural Network

In this Section, we first provide a Rocchio Nearest Centroid Laplacian Feature Selection model and then describe the Arbitrary Normalized Attention-based Recurrent Neural Network Lead Generation in detail. We then present the RNC-NNN method and corresponding implementations.

3.1. Problem Definition

Let us consider an undirected graph ' $G = (V, E)$ ' denoting the social network, where ' $V = \{U_1, U_2, \dots, U_n\}$ ' representing the user set and ' E ' the edge set associating them. Then, the edge

$(U_i, U_j) \in E$ denoting a social binding between U_i and U_j that in turn are referred to as users communicating at a time instance. We denote with TL the tweet log, a record of the tweets performed by every user in the social network. Each entry of the tweet log TL refers to a tuple (t_i, U_i, T_i) denoting the tweets T_i made by user U_i at time instance t_i . Let T be the set of tweets performed in TL , for each tweet $t \in T$, each user is either operative (if tweet action exist between two users) or non-operative (if tweet action does not exist between two users). We consider the scenario where features or tweets selected are found to be very closer to each other to taken into consideration the mutual coordination between the users while arriving at a decision-making process (selecting courses via digital marketing).

In this paper, we keep track of user tweets over time to model social relationships between them. In specific, we concentrate on the understanding and the designing of the digital marketing phenomenon, with the final objective of recommending the courses and universities list for the students in real-world scenarios for distance learning via lead generation. Once the method is trained, we target to use it to explore the tweets related to distance learning to understand people's opinion. More specifically, our objective is to infer whether the tweets related to distance learning will have an impact on the digital advertising.

To this aim, we introduce Rocchio Nearest Centroid and Normalized Neural Network (RNC-NNN), a deep learning method for modeling and forecasting digital marketing on the digital learning aspect. RNC-NNN is based on neural networks. First, relevant features or tweets are selected using Rocchio Nearest Centroid function. Next, the rationale of RNC-NNN is the potentiality of neural network to arbitrary extract complicated relationships for analyzing the polarities about distance learning. Thereby, if we denote each user's tweets in the social network as an input node for NNN we can model the interplay between user's tweets via the NNN layers. For each user's tweets, we consider the history of tweets, re-tweets made by them to train the NNN and fine tune the model parameters. Upon successful completion of NNN training, an Arbitrary Normalized Attention Mechanism is utilized to obtain lead generation resultant values.

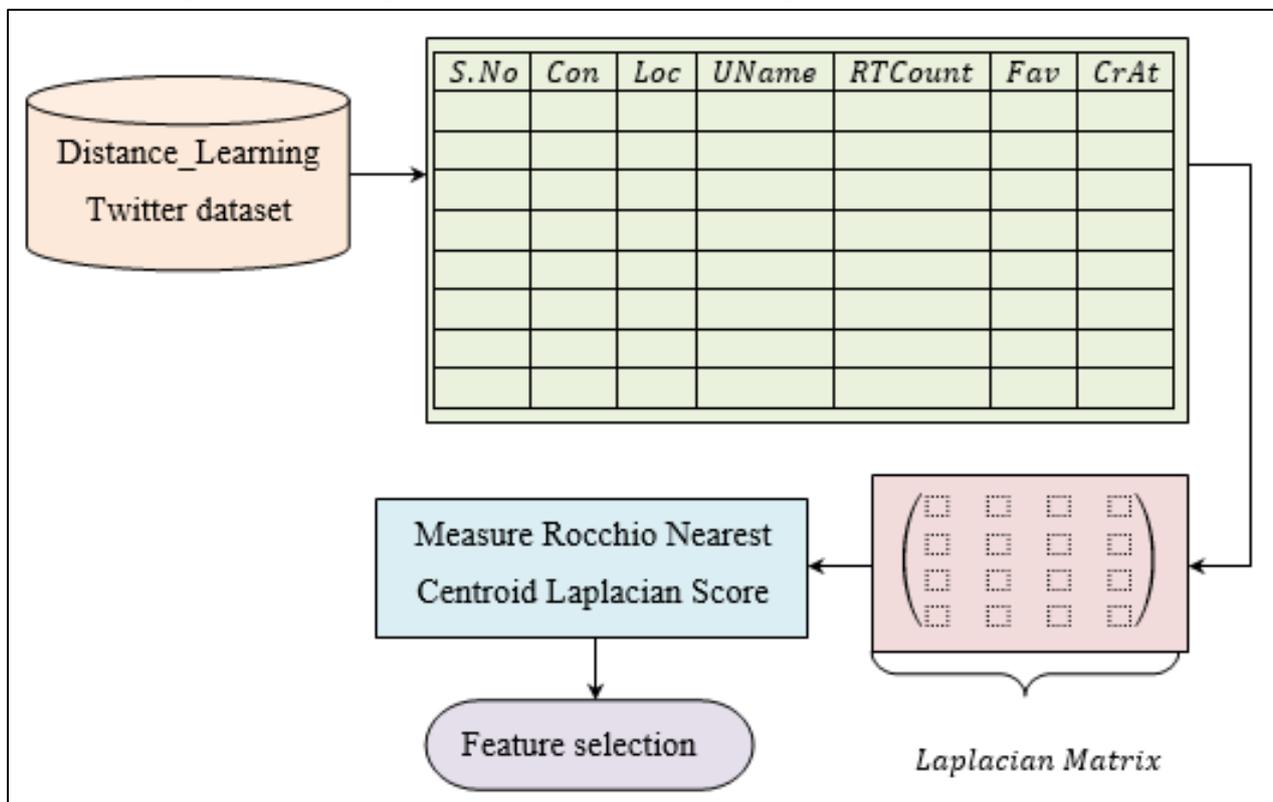
4.2. Rocchio Nearest Centroid Laplacian Feature Selection

Usage of social media (i.e., Twitter) has become an essential component for several users around the globe. Specifically, the impact of social media for the student community has changed the lives across the globe by selecting their choice of universities for education at their doorstep through tweets. With this the organizations have acknowledged to this alternate in studying community

behavior by making digital and social media an indispensable and fundamental element of their business marketing plans, therefore concentrating on digital marketing. However, with the presence of large numbers of tweets though all not be of use for further analysis, a significant amount of relevant features (i.e., tweets) has to be selected whereas the remaining insignificant features has to be discarded.

The principal purpose of feature selection remains in obtaining feature subset for data representation, so that those features with minimum relevancy are discarded and hence they are not considered for facilitating analysis. The main objective of feature selection hence remains in identifying a feature subset that results in high learning accuracy. Feature selection hence not only minimizes the data size and execution time of learning algorithms, but also results in better generalization capability. In this work, Rocchio Nearest Centroid Laplacian Feature Selection model is designed that not only selects the relevant features with minimum processing capability but also focuses on the accuracy aspect. Figure 1 shows the block diagram of Rocchio Nearest Centroid Laplacian Feature Selection model.

Figure 1 - Block Diagram of Rocchio Nearest Centroid Laplacian Feature Selection Model



As shown in the above figure, let us consider tweets on distance learning dataset [20], consisting of ‘n’ vectors, $\{T_i\}_{i=1}^n$ an adjacency matrix $W_{n \times n}$ between T_i and T_j data points, then ‘W’ denotes

a weighted graph, whose nodes represents the instances or users tweets made whereas the set of edges comprises of an association for each pair of nodes or users ' U_i ' with weight ' W_{ij} ' based on the Rocchio Nearest Centroid function. The Laplacian matrix ' LM ' is then mathematically represented as given below.

$$LM = DM - W \quad (1)$$

From the above equation (1), the results of the laplacian matrix ' LM ', is arrived at on the basis of the diagonal matrix ' DM ' such that ' $d_{ii} = \sum_{j=1}^n W_{ij}$ '. Then, given an undirected graph ' G ', the laplacian matrix ' LM ' on a vector ' $v \in R^n$ ', the local variation is formulated as given below.

$$v^T L v = \frac{1}{2} \sum_{i \neq j} W_{ij} (v_i - v_j)^2 \quad (2)$$

From the above equation (2) local variation quantification for each vector is measured. This local variation quantification motivates to utilize the laplacian matrix ' LM ' fact on a vector of values of feature concerning tweets along with the other features, the consistency of this feature regarding the structure of the undirected graph ' G '. Then, the feature (i.e., tweet) is said to be compatible with the structure of an undirected graph if it takes similar tweet values for instances that are closer to each other in the graph, and dissimilar tweet values for instances that are farther from each other. Thus a compatible feature or tweet would be relevant to separate the classes. If we denote the vector ' $v_i = (v_{i1}, v_{i2}, v_{i3}, \dots, v_{in})^T$ ' with ' $i = 1, 2, \dots, m$ ' as the ' i -th feature' and its value for the ' n ' instances, then the Rocchio Nearest Centroid Laplacian Score is mathematically formulated as given below.

$$LM_i = \frac{v_i^T L v_i}{v_i^T D v_i} \quad (3)$$

From the above equation (3), ' LM ' represents the laplacian matrix of an undirected graph ' G ', with ' D ' representing the diagonal matrix, ' v_i ' the variations from the average of all the considerations of the ' v ' vector. Followed by which, the per-class centroids are measured as given below.

$$v_i = v_i - \left[\frac{1}{|C_l|} \right] \sum_{i \in C_l} T_i \quad (4)$$

From the above equations (4) Rocchio Nearest Centroid functions is arrived at based on per-class centroids ' v_i ' with ' C_l ' denoting the indices set of tweet samples belonging to class ' $l \in F$ '. According to the Rocchio Nearest Centroid score, a good feature or tweet should possess a small value for ' LM_i '. Thus, the features are organized in a list in accordance to their relevancy. The features or tweets that are at the crest of the list are those with smaller values and hence are contemplated as the

most predominant. The pseudo code representation of Rocchio Nearest Centroid Laplacian feature selection is given below.

Algorithm 1 - Rocchio Nearest Centroid Laplacian Feature Selection

Input: Dataset ‘ DS ’, Features ‘ $F = F_1, F_2 \dots, F_n$ ’, Tweets ‘ $T = T_1, T_2, \dots, T_n$ ’
Output: Relevant feature (tweet) selection ‘ RT ’
1: Initialize undirected graph ‘ G ’, weight ‘ W ’, diagonal matrix ‘ DM ’ 2: Begin 3: For each dataset ‘ DS ’ with features ‘ F ’ and tweets ‘ T ’ 4: Formulate Laplacian matrix ‘ LM ’ as in equation (1) 5: Obtain local variation quantification for each vector as in equation (2) 6: Measure Rocchio Nearest Centroid Laplacian Score as in equation (3) 7: Measure per-class centroids as in equation (4) 8: Return relevant features (i.e., tweets) selection 9: End for 10: End

As given in the above algorithm, the objective here remains in selecting the relevant feature with better generalization capability. To achieve this objective, with distance learning dataset acquired as input, a laplacian matrix is formulated with local variation quantification, therefore reducing the processing time involved in selecting the feature. Next, with the aid of Rocchio Nearest Centroid Laplacian Score per-class centroids are measured hence retrieving the relevant feature and contributing to accuracy.

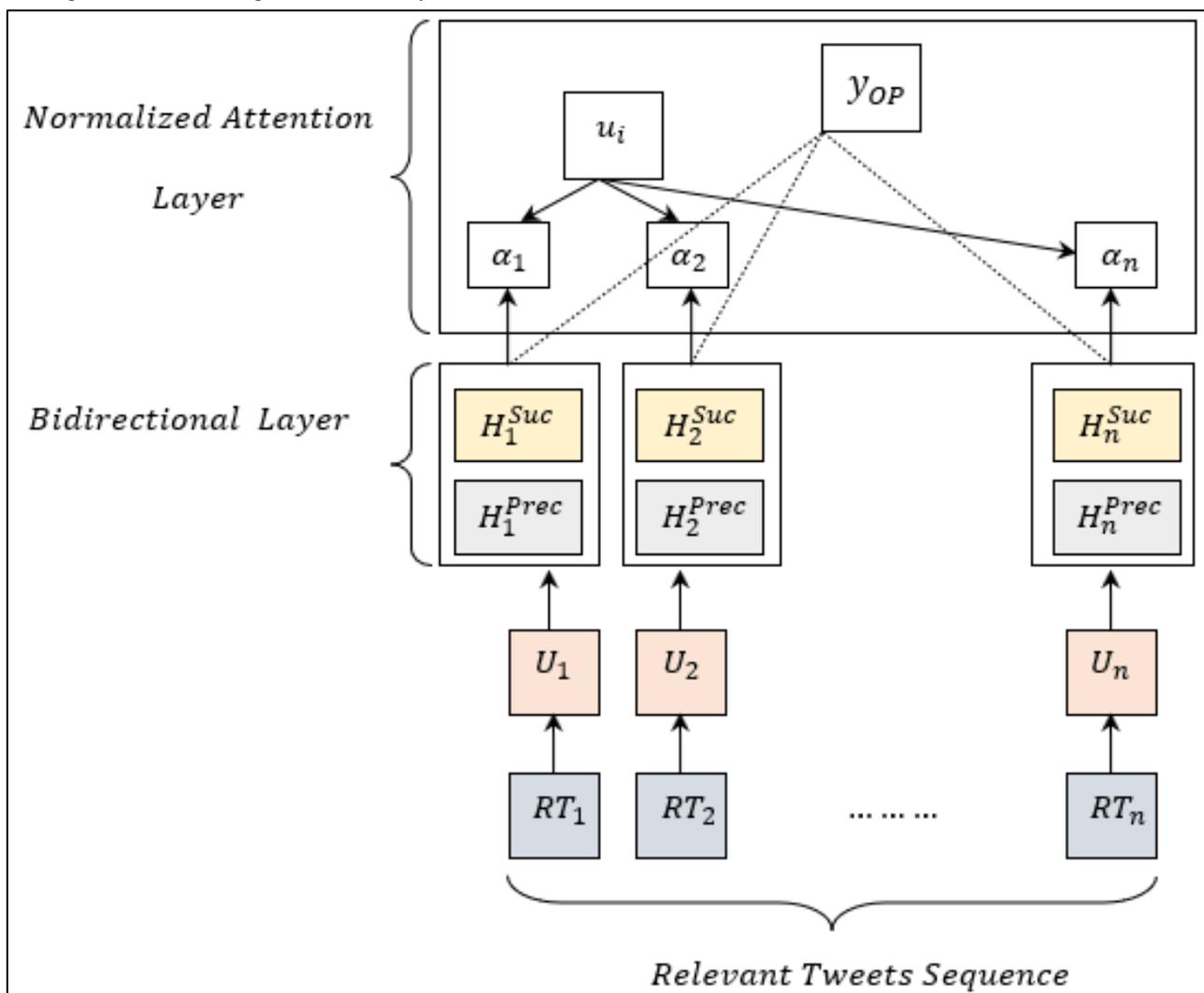
4.3. Arbitrary Normalized Attention-based Recurrent Neural Network Lead Generation

In this section, with the relevant features selected, a deep learning-based lead generation model using Arbitrary Normalized Attention-based Recurrent Neural Network Lead Generation is proposed. Let us consider relevant tweets ‘ $RT = [RT_1, RT_2, \dots, RT_n]$ ’, where the relevant tweets ‘ $RT_i \in RT_{OP}$ ’ are in chronological order, ‘ OP ’ is the set of relevant online opinions for the corresponding relevant tweets ‘ RT ’. With the objective of generating the lead with maximum precision, our Arbitrary Normalized Attention-based Recurrent Neural Network Lead Generation model measures the following probability.

$$Prob(conv\ on\ OP|RT) = Prob(y_{OP} = 1|RT) \quad (5)$$

From the above equation (5), ' y_{OP} ' represents the binary variable denoting conversion ' $(y_{OP} = 1)$ ', for positive tweets or where ' $(y_{OP} = 0)$ ', for negative tweets respectively. Figure 2 shows the block diagram of Arbitrary Normalized Attention-based Recurrent Neural Network Lead Generation model.

Figure 2 - Block Diagram of Arbitrary Normalized Attention-based Recurrent Neural Network Lead Generation



As shown in the above figure, relevant tweets embedding layer, bidirectional layer, and the normalized attention layer are the paramount elements of the model. To start with the relevant tweet ' $RT = [RT_1, RT_2, \dots, RT_n]$ ' is fed as input to the relevant tweets embedding block, which learns a polarity representation ' Pol_i ' for each tweet in ' $RT_i \in RT_{OP}$ '. The arbitrarily initialized embeddings are learnt as a part of the Normalized Attention-based Recurrent Neural Network Lead Generation model training. Next, as illustrated in the above figure, the user tweet embedding sequence

' $[U = U_1, U_2, \dots, U_n]$ ' is fed to a bidirectional layer. The final output representation is acquired by aggregating characterizations from the preceding '*Prec*' and succeeding '*Suc*' cell. This is mathematically represented as given below.

$$H_i = [H_i^{(Prec)}, H_i^{(Suc)}] \quad (6)$$

From the above equation (6), bidirectional layer symbolizes the relevant tweets of the respective users in the form of hidden representation in both the preceding ' $H_i^{(Prec)}$ ', and succeeding ' $H_i^{(Suc)}$ ', states respectively. The conventional attention mechanism, an attention layer is added on top of the RNN module, in order to differentiate the handout of each output from the succeeding layer towards final lead generation. In specific, it initially transforms the RNN outputs ' H_i ' to a normalized dimensional representation ' u_i ' as given below.

$$u_i = \tanh(W_j H_i + b_j) \quad (7)$$

From the above equation (7), the normalized dimension is obtained based on the weight ' W_j ', bias ' b_j ' and the hidden representation (i.e., covering both the preceding and succeeding states) ' H_i ' respectively. Followed by which a context vector ' u_j ' is introduced to obtain the tweet polarities, during the RNN training process. In other words, it evaluates how much normalized attention should be given to each input representation ' u_i '. The normalized attention layer here estimates the interior product between ' u_i ' and ' u_j ' and finally normalizes with the aid of the softmax function as given below.

$$\alpha_i = \frac{\exp(u_i^T u_j)}{\sum_{i=1}^n \exp(u_i^T u_j)} \quad (8)$$

From the above equation (8), the normalized weight ' α_i ' is obtained based on each input representation ' u_i ' and context vector ' u_j ' respectively. The attention layer outputs ' y_{OP} ' that represents the aggregated summation value of all the normalized dimensional vectors ' u_i ' of each relevant user tweet for producing the final lead as given below.

$$y_{OP} = \sum_{i=1}^n \alpha_i H_i \quad (9)$$

Since this attention mechanism only estimates the attention scores in terms of normalized trails in an arbitrary manner, we name this attention model as Arbitrary Normalized Attention Mechanism (ANAM), i.e., it has attention arbitrarily normalized to a trail. The pseudo code representation of Arbitrary Normalized Recurrent Neural Network Lead Generation is given below.

Algorithm 2 - Arbitrary Normalized Recurrent Neural Network Lead Generation

Input: Dataset ‘ DS ’, Features ‘ $F = F_1, F_2 \dots, F_n$ ’, Tweets ‘ $T = T_1, T_2, \dots, T_n$ ’
Output: Precise lead generation
<pre> 1: Initialize tweet polarities ‘u_j’ 2: Begin 3: For each dataset ‘DS’ with features ‘F’ and relevant tweets ‘RT’ 4: Formulate binary variable probability as in equation (5) //input layer 5: Obtain relevant tweets //bidirectional layer 6: Feed user tweet embedding sequence to bidirectional layer as in equation (6) //normalized attention 7: Obtain normalized dimensional representation as in equation (7) 8: Estimate normalized weight as in equation (8) 9: Evaluate Arbitrary Normalized Attention as in equation (9) 10: If ‘$y_{OP} = 1$’ 11: Relevant tweets indicate positive tweets 12: End if 13: If ‘$y_{OP} = 0$’ 14: Relevant tweets indicate negative tweets 15: End if 16: Return (score, label) 17: End for 18: End </pre>

As given in the above algorithm, the objective remains in generating the lead with high precision or positive lead results. To achieve this aim, a deep learning model using Arbitrary Normalized Recurrent Neural Network is applied in our work. First, the relevant tweets are provided as input to the input layer, followed by which bidirectional layer is formalized to create a two-way direction between the preceding and succeeding cell. Finally, normalized attention function is applied to evaluate the polarities with which binary variable symbolizing positive tweets or negative tweets are arrived at. With this tweet results, lead are said to be generated for education as a domain where the students can make a clear decision to determine tweet polarities, related to distance learning.

4. Experimental Settings

In this section, we report experiments and comparisons made with the existing state-of-the-art methods Deep Learning-based frameworks [1] and ML integrated social media marketing (ML-SMM)

[2] on tweets collected from distance learning dataset to analyze our proposed Rocchio Nearest Centroid and Normalized Neural Network (RNC-NNN) method. Experimental evaluations are performed in Python by utilizing distance learning [20] employing three distinct csv files, i.e., raw files, processed files and sentiment files.

5. Discussion

Comparative analysis of lead generation methods is performed and compared with three different methods RNC-NNN, Deep Learning-based frameworks [1] and ML-SMM [2]. Performance analysis is made with three distinct parameters namely, processing time, lead generation accuracy and precision for the respective number of tweets and distinct tweet sizes.

5.1. Scenario 1: Processing Time

In this section, the processing time analysis is made with the aid of distance learning dataset. For simulation, the user count in the range of 500 to 5000 was used with the tweet size in the range of 1000 to 10000 with the tweets gathered from different users, location, time and date. The processing time analysis was made here to estimate the time consumed in obtaining relevant tweets. This is mathematically expressed as given below.

$$PT = \sum_{i=1}^n UC_i * Time [RT] \quad (10)$$

From the above equation (10), the processing time ‘*PT*’, is measured based on the user count ‘*UC_i*’ taken into consideration and the time consumed in acquiring relevant tweet ‘*Time [RT]*’. It is measured in terms of milliseconds (ms).

Table 1 - Processing Time Results

User_count	Processing time (ms)		
	RNC-NNN	Deep Learning-based frameworks	ML-SMM
500	22.5	29	35
1000	31.35	45.35	62.35
1500	40.25	51.25	75.55
2000	48.35	63.25	90.35
2500	59.15	75.45	105.25
3000	70.25	93.15	125.35
3500	73.55	103.55	140.55
4000	95.15	125.55	165.35
4500	105.25	140.35	185.35
5000	135.55	180.15	200.25

Table 1 reports the result of processing time using RNC-NNN, Deep Learning-based frameworks [1] and ML-SMM [2]. The initial apparent result is that deep learning method outperforms the state-of-the-art methods which verify the power of recurrent neural networks to capture linear correlation between tweets and classes. Comparing three different methods based on processing time, our proposed RNC-NNN method is the best among all. As our proposed method pays more attention to feature subset for data representation, having higher performance in our proposed RNC-NNN method selects the relevant feature for lead generation with minimum processing time.

Figure 3 - Graphical Representation of Processing Time

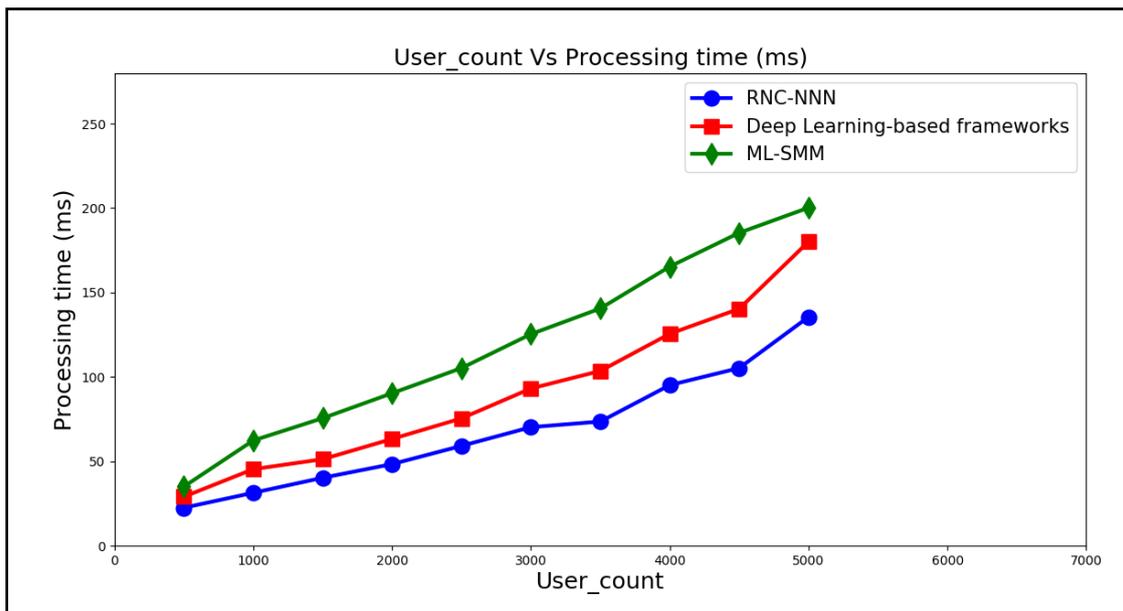


Figure 3 shows the processing time for different numbers of user counts ranging from 500 to 5000. As depicted in the above figure, processing time is directly proportional to the user count considered for simulation. The blue color line denotes the processing time of the proposed RNC-NNN method, the red color line denotes the processing time of energy consumption of Deep Learning-based frameworks [1] and green color line denotes the processing time of ML-SMM [2] respectively. The results shows that the processing time of RNC-NNN method is minimal than [1] and [2]. This is because of applying the Rocchio Nearest Centroid Laplacian Feature Selection model to select the relevant features taking into consideration the generalization capability. After that, Rocchio Nearest Centroid Laplacian Score is used to determine the closest tweets. The tweets that are found to be closer with each other are selected for determining per-class centroids. Consequently, processing time using RNC-NNN method is found to be reduced by 25% compared to [1] and 43% compared to [2].

5.2. Scenario 2: Lead Generation Accuracy

The second parameter of consideration taken into account this section is the lead generation accuracy. For simulation purpose, tweet size in the range of 250 to 2500 related to distance learning to apprehend people's opinion are collected from 5000 different users. Also the tweet length varied in the range of 25 to 100, also taking into consideration the retweet count values for acquiring the lead. With all this factors into consideration, the lead generation accuracy is measured as given below.

$$LGen_{acc} = \sum \frac{TRet_{acc}}{T_{size}} * 100 \quad (11)$$

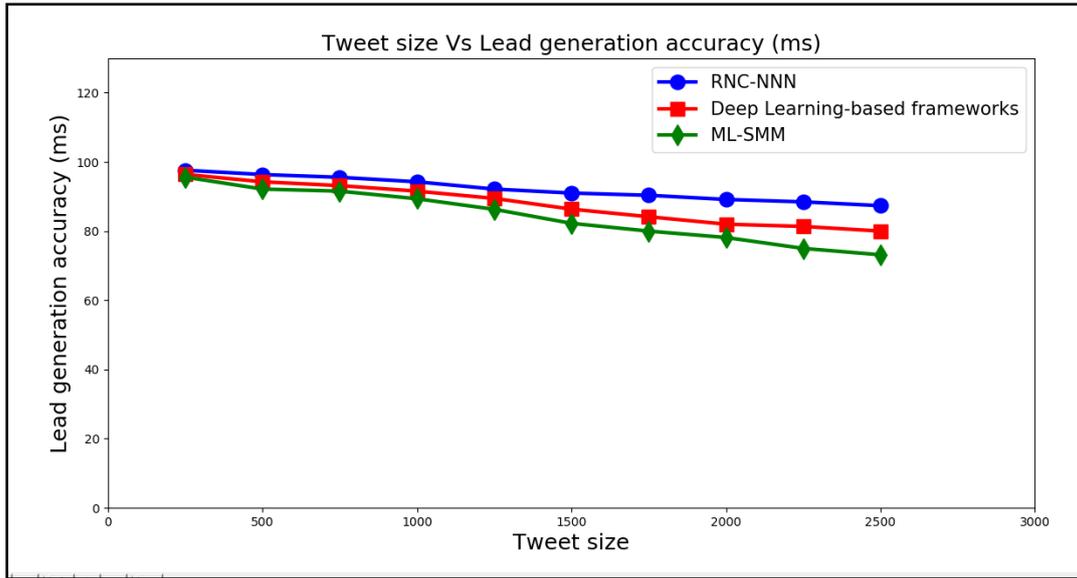
From the above equation (11), the lead generation accuracy ' $LGen_{acc}$ ' is measured based on the tweet size ' T_{size} ' and the tweets retrieved accurately ' $TRet_{acc}$ '. It is measured in terms of percentage (%).

Table 2 - Lead Generation Accuracy Results

Tweet size	Lead generation accuracy (ms)		
	RNC-NNN	Deep Learning-based frameworks	ML-SMM
250	97.6	96.4	95.6
500	96.35	94.25	92.15
750	95.55	93.15	91.55
1000	94.25	91.55	89.35
1250	92.15	89.45	86.35
1500	91	86.35	82.25
1750	90.35	84.15	80
2000	89.15	82	78.15
2250	88.45	81.35	75
2500	87.35	80	73.15

Table 2 also summarizes the performance of different methods for lead generation accuracy using other performance metrics. Considering the lead generation accuracy values, the overall results in Table 2 illustrate that the proposed method in this social learning task for distance learning outperforms the others. It shows the effectiveness of Rocchio Nearest Centroid Laplacian Score per-class centroids in generating the lead in an accurate manner. Having the higher performance for neural network methods upon comparison to the state-of-the-methods like [1] and [2] highlights the significance of per-class centroids in indices set of tweet samples.

Figure 4 - Graphical Representation of Lead Generation Accuracy



In figure 4, we depict the lead generation accuracy of our method by making elaborate comparisons with [1] and [2] at varying tweet size. Different aspects are significance of inspection. There is a discernible aperture between the performance of [1] and [2] upon comparison with our proposed method, RNC-NNN. This is likely due to higher involvement of tweet polarities in a social event while exploring what to be contemplated about distance learning. Further, we can infer that our method, on median, is apt to accurately retrieve the tweets about 97.6% of user’s tweets in case of tweet size being 250, 96.4% of user’s tweets using [1] and 95.6% of user’s tweets using [2]. This result further highlights the weakness of accuracy using long short term memory in social media marketing with only, 244 tweets accurately retrieved using the proposed RNC-NNN method, 241 tweets using [1] and 239 tweets using [2] respectively. Such findings suggest that the lead generation accuracy using RNC-NNN method was observed to be higher by 5% compared to [1] and 10% compared to [2].

5.3. Scenario 3: Precision

Finally, precision is measured in this section. Precision also known as the positive predictive value refers to the ratio of relevant leads among the retrieved leads. This is measured as given below.

$$P = \frac{RelLead_{ret}}{Lead_{ret}} * 100 \tag{12}$$

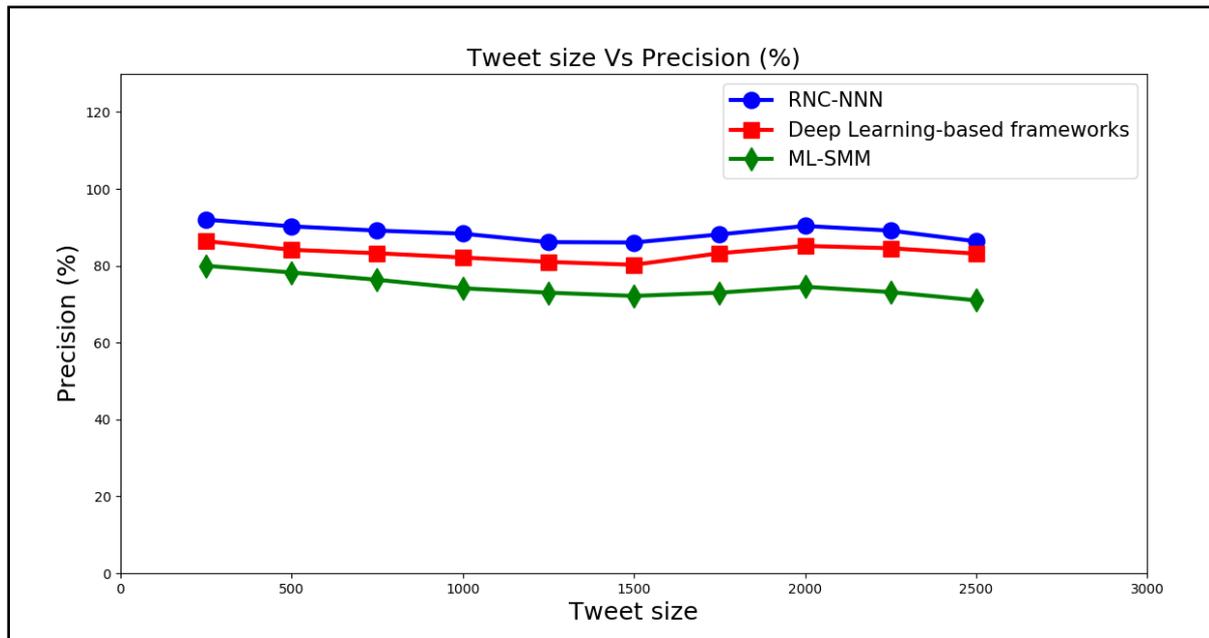
From the above equation (12), precision ‘P’, is measured on the basis of the number of relevant leads retrieved ‘RelLead_{ret}’ to the number of leads retrieved ‘Lead_{ret}’. It is measured in terms of percentage (%).

Table 3 - Precision Results

Tweet size	Precision (%)		
	RNC-NNN	Deep Learning-based frameworks	ML-SMM
250	92	86.4	80
500	90.25	84.15	78.25
750	89.15	83.25	76.35
1000	88.35	82.15	74.15
1250	86.15	81	73
1500	86.05	80.25	72.15
1750	88.15	83.25	73
2000	90.35	85.15	74.55
2250	89.15	84.55	73.15
2500	86.35	83.15	71

It is worth noting that in Table 3, our method (RNC-NNN) show the best precision result compared to other state-of-the-art methods for the same tweet size represented in Table 2. This is specifically owing to the fact that our objective remains in maximizing the lead generation and polarities in terms of the precision score, which combines the polarity accuracy on both positivity and negativity of the tweets. As polarity focuses on large portion of the dataset, precision reflect the genuine performance of the classifier for lead generation.

Figure 5 - Graphical Representation of Precision



Finally, figure 5 given above shows the graphical representation of precision with respect to three different methods, RNC-NNN, Deep Learning-based frameworks [1] and ML-SMM [2]. From the above figure it is observed that the precision rate is neither found to be increased nor found to be decreased with respect to distinct tweet sizes in the range of 250 to 2500. This is owing to the reason that increasing the tweet size acquired from the distinct users also results in the significant increase in the retweet count value and this in turn causes an increase or decrease in value in the precision rate. However, simulations performed for tweet size of 250, leads retrieved was 125 for all the three methods, whereas number of relevant leads retrieved using the three methods were found to be 115, 108 and 100 respectively. With this the overall precision were observed to be 92%, 86.4% and 80%. The reason behind the improvement was owing to the application of Arbitrary Normalized Recurrent Neural Network Lead Generation algorithm. By applying this algorithm, relevant tweets were given as input, next, bidirectional layer was formalized to obtain a two-way direction between the preceding and succeeding cell. Finally, normalized attention function was then applied to measure the polarities. With this the precision rate using RNC-NNN was said to be improved by 6% compared to [1] and 19% compared to [2] respectively.

6. Conclusion

This paper presents a lead generation method called, Rocchio Nearest Centroid and Normalized Neural Network in social media marketing for precision performance enhancement with minimum processing time and maximum lead generation accuracy. The main objective of the proposed RNC-NNN method is to select the most relevant features or tweets for performing efficient lead generation. The laplacian matrix, Rocchio Nearest Centroid Laplacian Score and per-class centroids are computed for relevant feature selection with better generalization capability. Relevant features or tweets are obtained via per-class centroids. Next, Arbitrary Normalized Attention-based Recurrent Neural Network Lead Generation model generates the lead depending on Arbitrary Normalized Attention Mechanism (ANAM). The arbitrary normalized attention value therefore forms two-way direction between the preceding and succeeding cell, ensuring marketing tweet funnel. The polarity values are obtained and finally, the leads are generated with which the positive and negative tweets are determined for analyzing the feedback. The aggregated leads are then provided to the user or the student community for selecting courses in universities with higher lead generation accuracy, precision and lesser processing time. Simulation results demonstrate that RNC-NNN method reduces the processing time involved in tweet and lead generation with higher precision and lead generation accuracy.

References

- Zhabiz Gharibshah, Xingquan Zhu, Arthur Hainline, Michael Conway, “Deep Learning for User Interest and Response Prediction in Online Display Advertising”, *Data Science and Engineering*, Springer, Jan 2020 [Deep Learning-based frameworks]
- B. Senthil Arasu, B. Jonath Backia Seelan, N. Thamaraiselvan, “A machine learning-based approach to enhancing social media marketing”, *Computers and Electrical Engineering*, Elsevier, May 2020 [ML integrated social media marketing (ML-SMM)]
- Thomas Davenport, Abhijit Guha, Dhruv Grewal, Timna Bressgott, “How artificial intelligence will change the future of marketing”, *Journal of the Academy of Marketing Science*, Springer, Feb 2020.
- Iuliana Mihaela Lazar, Georgeta Panisoara, Ion Ovidiu Panisoara, “Digital technology adoption scale in the blended learning context in higher education: Development, validation and testing of a specific tool”, *PLOS ONE*. <https://doi.org/10.1371/journal.pone.0235957> 2020
- Yogesh K. Dwivedi, Elvira Ismagilova, D. Laurie Hughes, Jamie Carlson, Raffaele Filieri, Jenna Jacobson, Varsha Jain, Heikki Karjaluo, Hajer Kefi, Anjala S. Krishen, Vikram Kumar, Mohammad M. Rahman, Ramakrishnan Raman, Philipp A. Rauschnabel, Jennifer Rowley, Jari Salo, Gina A. Tran, Yichuan Wang, “Setting the future of digital and social media marketing research: Perspectives and research propositions”, *International Journal of Information Management*, Elsevier, 2021.
- Michal Podhoranyi, “A comprehensive social media data processing and analytics architecture by using big data platforms: a case study of twitter flood-risk messages”, *Earth Science Informatics*, Springer, 2021.
- Ayesha Rashid, Muhammad Shoaib Farooq, Adnan Abid, Tariq Umer, Ali Kashif Bashir, Yousaf Bin Zikria, “Social media intention mining for sustainable information systems: categories, taxonomy, datasets and challenges”, *Complex & Intelligent Systems*, Springer, Mar 2021
- Md Rafiqul Islam, Shaowu Liu, Xianzhi Wang, Guandong Xu, “Deep learning for misinformation detection on online social networks: a survey and new perspectives”, *Social Network Analysis and Mining*, Springer, 2020.
- Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang, “FakeBERT: Fake news detection in socialmedia with a BERT-based deep learning approach”, *Multimedia Tools and Applications*, Springer, Jan 2021.
- Ishaani Priyadarshini, Chase Cotton, “A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis”, *The Journal of Supercomputing*, Springer, May 2021.
- Malik Khizar Hayat, Ali Daud, Abdulrahman A. Alshdadi, Ameen Banjar, Rabeeh Ayaz Abbasi, Yukun Bao, Hussain Dawood, “Towards Deep Learning Prospects: Insights for Social Media Analytics”, *IEEE Access*, Apr 2019.
- Samik Datta, Satyajit Chakrabarti, “Aspect based sentiment analysis for demonetization tweets by optimized recurrent neural network using fire fly-oriented multi-verse optimizer”, *Sadhana*, Mar 2021
- Massa Baali and Nada Ghneim, “Emotion analysis of Arabic tweets using deep learning approach”, *Journal of Big Data*, Springer, Feb 2019.
- Ming-Hui Huang, Roland T. Rust, “A strategic framework for artificial intelligence in marketing”, *Journal of the Academy of Marketing Science*, Springer, Nov 2020.

Liang-Chu Chen, Chia-Meng Lee, Mu-Yen Chen, “Exploration of social media for sentiment analysis using deep learning”, *Soft Computing, Springer*, Oct 2019.

Shiwei Liu, Iftitahu Ni'mah, Vlado Menkovski, Decebal Constantin Mocanu, Mykola Pechenizkiy, “Efficient and effective training of sparse recurrent neural networks”, *Neural Computing and Applications, Springer*, Jan 2021

Kamila Migdał-Najman, Krzysztof Najman, Sylwia Badowska, “The GNG neural network in analyzing consumer behavior patterns: empirical research on a purchasing behavior processes realized by the elderly consumers”, *Advances in Data Analysis and Classification, Springer*, Aug 2020

György Kovács, Pedro Alonso, Rajkumar Saini, “Challenges of Hate Speech Detection in Social Media”, *Computer Science, Springer Nature*, Feb 2021.

Clemens Koob, “Determinants of content marketing effectiveness: Conceptual framework and empirical findings from a managerial perspective”, *PLOS ONE*.

<https://doi.org/10.1371/journal.pone.0249457> April 1, 2021

https://github.com/Bhasfe/distance_learning