*International Journal of Finance, Accounting and Economics Studies, 3(4), 2022, 33-46.*
*Print ISSN: 2251-645X Online ISSN: 2383-2517*

**Islamic Azad University Science and Research Branch**
**Faculty of Management and Economics**

**International Journal of Finance, Accounting and Economics Studies**
**Journal homepage: https://ijfaes.srbiau.ac.ir**


# Prediction of Message Diffusion: A Deep Learning Approach on Social Networks

**Hosniyeh Safiarian [1], Mohammad Jafar Tarokh [2*], Mohammad Ali Afshar Kazemi [3]**

*[1] P.hd Student, Faculty of Management and Economic, Science & Research Branch, Islamic Azad University, Tehran, Iran. ho.safiarian@gmail.com*
*[2*] Professor, IT Group ,Faculty of Industrial Engineering, K. N. Toosi University of Technology Tehran, Iran. Mjtarokh@kntu.ac.ir*
*[3] Associate Professor, Management Group, Faculty of Management ,Tehran North Branch, Islamic Azad University, Tehran, Iran. Dr.mafshar@gmail.com*

**Article History**
**Submission Date:** 01 January 2022
**Revised Date:** 25 February 2022
**Accepted Date:** 29 March 2022
**Available Online:** December 2022

**JEL Classification:**

**Abstract**
Nowadays, many industries pay attention to social media because people are spending sizable chunks of their lives in virtual worlds. Some of the social networks such as Facebook, Instagram and Twitter affected by their user through content. Predicting the popularity of content can play an important role in different areas such as viral marketing, advertising and propagation news. However, prediction problem is a challenging problem. In this paper, we developed a deep learning approach to predict the popularity of tweets in the twitter social network. It is called DLMD. We extracted the feature of content from each tweet. We use the deep learning approach for prediction it means that we model this problem with a binary classification problem Our proposed method evaluate with different measures and the results show that DLMD method has a high accuracy in prediction rather than other methods. Therefore, DLMD is a convenient method to predict diffusion on social network.

**Keyword:**
Deep learning
Influential Content
Social Network
Prediction, Users

* Corresponding Author: Mjtarokh@kntu.ac.ir

## 1. Introduction

Social media is such a suitable method to find information about different problems. Due to popularity of the Online Social Network (OSN), a large number of people are now utilizing OSN services in order to gain active collaboration, participation and interaction with other users. Therefore , social media can open new opportunities for different industries.

Microblogging is a content-oriented concept in which people can interact with others both known and unknown (Ni et al., 2021). Users can react rather than content. They can share contents easily with other users. Then, modeling and predicting the popularity of online contents has become an important research topic in social media analytics and is beneficial to many applications in public management, business and security-related domains (Fang and Aili, 2021). However, Users with many connections can suffer from information overload. It is quite important to filter information flow for the end users and to provide them with important content (Sandberg, John, et al., 2019).

Popularity prediction is also helpful in personalizing the content and finding the right tweets for end users. On the other hand, understanding how and why a tweet becomes popular, can help to gain a better insight into how the information is dispersed over the network (Li et al., 2016). According to previous studies, Feature-based methods are very famous for prediction (P. Cerchiello et al., 2017). These methods extracted a large number of features for example the structure of networks, users, contents and then train Machin learning models to predict popularity. They are good methods, however, these methods have something to do with feature engineering. Some researchers show that timeserie is a suitable method for prediction but this method ignore other valuable information for popularity prediction (S. Mishra et al., 2016). Some studies show that deep learning can have an important role in prediction, however, these methods consider only user information (C. Li et al., 2017) .

In this paper, we extracted the features of content, These features have direct effect on diffusion ,on the other hand they can play a significant role in diffusion . We make a proposed method for prediction by a set of binaries classify and deep learning approach can be useful for it. It is called DLMD. Also, we evaluated the proposed method with other classifiers methods.

This paper is organized as follows: in section 2 we discuss the work related to our research. Proposed method is explained in section 3. In section 4 we explain, in some detail. The datasets and experiments that we have conducted. Finally, we tell some conclusions and discuss the possible future directions in section 5.

## 2. Related work

In this section, we provide an overview of the existing approaches to popularity prediction in social networks.

### 2.1 Epidemics Models in social networks

Information diffusion is similar to the spread of an epidemic, but there are differences. Information diffusion on the social network has time, relationship strength, information content, social factors, network structure (Liu et al., 2014), etc. In the context of epidemic propagation analysis, one of the famous propagation models and widely used models is the susceptible– infected–recovered (SIR) model, which is composed of three categories of individuals (Anderson, 1992). In (Wang et al., 2014) developed the SEIR model (Liu et al., 2014) by inserting Exposed (E) nodes based on SIR model. They made a dynamical evolution equation to describe the information diffusion, analyzing the impact of user logging frequency and number of friends on information diffusion. John and Joshua (Cannarella et al., 2014) are developed the SIR model and they proposed irSIR model to simulate the adoption and abandonment of user views by adding an infection recovery kinetics process. In (Feng et al., 2015) proposed a FSIR model. The model has something to do with the neighbors that they have a positive impact on an individual in the diffusion of information. In (Wang et al., 2015) developed SIS model and proposed ESIS model based on emotional information. It proved that information diffusion is related to propagation probability and transmission intensity, and therefore its performance is better than that of the SIS model

**2.2 influence maximization**

Influence maximization is a recent focus of social networks research's-authorship data studied viral marketing problem using several commonly used diffusion models such as the Linear Threshold, Independent Cascade and General Threshold Models. They discover influence maximization problem is NP hard and also presented a greedy algorithm for influence maximization problem (Kempe et al., 2003). Large water distribution network focused on the scalability of greedy approach of influence maximization ,this algorithm is called CELF (leakovec et al., 2007). Two real life collaboration networks are studied to reduce running time of the greedy algorithm of (Kempe et al., 2003) .and it is improved by (leakovec et al., 2007), this algorithm is called Mix greedy (Chen et al., 2010) .In (Mishra, 2016). Hawkes processes is designed to model the spreading processes of content. They pay attention to two important factors : learned stochastic processes parameters and some other features to train a random forest model (Mishra, 2016). In (Li, 2017) a cascade network is created by nodes as users and relationships as edged. It takes random walks on the cascade network. Also , features are extracted from the walk sequences by Neural Network (RNN) for popularity prediction. In (Trzciński, 2017) features are extracted from video content and features use to predict the number pf views at a future time. The method is limited to video popularity prediction. In(Cao et al., 2017) Hawkes method (Mishra, 2016) is extended and is called DeepHawkes. It uses user embedding vector and adopts RNN to encode cascade path. In (Hong, 2011) is the first to predict the popularity of text messages in microblogs. This is a kind of project that they focus on the number of retweets of a message in Twitter. They categorize the popularity values to different levels and consider each message as an independent instance for prediction. In (Yang et al., 2010) They assume that information can be diffused between users on the social networks and proposed method is defined based on speed, scale and range of diffusion. In (Vallet et al., 2015) pay attention to relation between textual features of tweets and view counts for YouTube Videos and propose a cross-system popularity prediction. In (Stowe et al., 2018) focus on the retweeting behavior by users in the geographically vulnerable areas and find out that users in the path of disasters prefer the diffusion of locally-created tweets over general information. In (Wang et al., 2019) proposed an efficient method to identify influential users in dynamic networks by exploiting local detection and updating strategy. In (Bhowmick et al., 2019) The goal of authors is to detect a set of seed nodes that maximized the influence propagation at the end of the diffusion process based on temporal retweet sequence between two consecutive retweet events .

## 3. Proposed method

Examining the retweeting behavior of users at the individual level can provide important information about the dissemination of information at the community level.

Tweet $\boldsymbol{tw_i}$ is effective for the user $\boldsymbol{u_i}$ if and only if it leads to a retweeting behavior $\boldsymbol{RT_{tw_i}}$ by the user $\boldsymbol{u_i}$. Measuring the effectiveness of a tweet requires the growth of a model for predicting the individual behavior of users in facing different tweets. Today, the function of social networks has gone beyond entertainment and these networks also operate with applications such as education, information and problem solving. Users often follow related pages to access the latest news and information in their area of interest. According to these functions, in this paper, a method based on the concept of content value of each tweet is used to investigate the retweeting behavior of users. It should be noted that content value is a relative concept and varies for different users. Depending on the field of activity and the amount of information and expertise, a tweet can have different content value for different users. To evaluate the content value of each tweet according to the retweeting behavior of users, in this study, a content-based method is presented.

**Definition 1**: Retweeting behavior: If a microblog is considered as a graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$, V is a set of users and E is a set of communications as a following. The following relationship between two users $\boldsymbol{U_i}, \boldsymbol{U_j}$as $\boldsymbol{U_i} \rightarrow \boldsymbol{U_j}$ indicates the following of user j by user i. Retweet behavior is represented as a triad $\boldsymbol{U_j}, \boldsymbol{tw_i}, \boldsymbol{t}$ , which means retweeting the tweet belonging to user $\boldsymbol{U_j}$, at time t. Each user $\boldsymbol{U_i}$has two sets of retweeted tweets $\boldsymbol{RP_{u_i}}$ and a list of posted tweets $\boldsymbol{p_{u_i}}$. The retweeting behavior of user $\boldsymbol{U_i}$ when viewing a tweet $\boldsymbol{tw_i}$ is displayed as a binary value $y(tw_i, u_i) = \{0,1\}$. $\boldsymbol{y(tw_i, U_i) = 0}$ indicates that tweet $\boldsymbol{tw_i}$ was not retweeted by $\boldsymbol{U_i}$, and $\boldsymbol{y(tw_i, U_i) = 1}$ indicates that $\boldsymbol{tw_i}$ tweet was retweeted by $\boldsymbol{U_i}$.

**Definition 2:** Problem of retweeting behavior prediction- For each tweet $\boldsymbol{tw_i} \in \boldsymbol{P}$ (P, set of tweets), $\boldsymbol{T} \in \boldsymbol{t}$ (T, time) and each user $u_i \in U$ (U, set of users) there is a function $\delta_t(tw, u_i)$ whose value is 1 if and only if a behavior (event) of retweeting $\boldsymbol{tw_i}$by user $\boldsymbol{U_i}$ at time $t' > t$ is occurred, otherwise the value of this function is zero. The problem of predicting the retweeting behavior of the user in the viewing a tweet is predicting the value of this function for $\forall tw \in P$and $\forall u_i \in U$.

### 3.1 Measuring the content value of a tweet

There are several informal definitions of content value in various sources. In this study, the content value of a tweet refers to the amount of attention that a tweet brings to its audience. This attention is shown in the form of retweeting, replying or mentioning the event by the user. As mentioned, for this end, the proposed method is based on content-based measurement and tweet statistics, which are discussed below.

#### 3.1.1 Checking the novelty of a tweet

For examining the content of a tweet in order to measure the content value, a lot of work has been done in the field of news. Many features are provided for reviewing the content, including a series of features related to the message (length, questionable or not, etc.) , tweeter (number of followers, year of registration, etc.) , topic of the message.

These methods perform classification, thematic ranking and similarity measurements by lexical analysis of words and statistical techniques. In the proposed method, new features are considered for online measurement of the value of information based on content and function of the problem. The novelty of a tweet is an important feature in its content value. By definition, a tweet $\boldsymbol{tw_i}$ is new to the user $\boldsymbol{U_i}$ if it is related to interests of $\boldsymbol{U_i}$ and unknown to him. Methods are mentioned for calculating unexpected information of web page retrieval in (Wu et al., 2011) and for unexpected news in (Alshaabi et al., 2020; Hansen et al., 2011).

$\boldsymbol{p_{u_i}} = \{\boldsymbol{tw_1}, \boldsymbol{tw_2}, \ldots. \boldsymbol{tw_n}\}$ is the set of tweets created by user $\boldsymbol{U_i}$. As stated in the definition of a tweet, the tweet must first be related to the user's interests. In order to determine the relationship of a tweet with the areas of interest to the user, in most of the work done, the TF-IDF method and the cosine of the angle between the words vector $\boldsymbol{tw_i}$ and $\boldsymbol{p_{u_i}}$ are used. This method has low accuracy due to the variety of words used. For this purpose, using the Latent Dirichlet Allocation topic modeling method, different topics in the texts can be categorized. Each topic contains a set of words $\boldsymbol{M_{topic\#}} = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots. \boldsymbol{w_n}\}$ ,the probability of which is specified in the relevant topic. To increase the accuracy of measuring the relationship between tweet $\boldsymbol{tw_i}$ and the set of tweets created by user $\boldsymbol{U_i}$, $\boldsymbol{p_{u_i}} = \{\boldsymbol{tw_1}, \boldsymbol{tw_2}, \ldots. \boldsymbol{tw_n}\}$, thematic similarity is measured using equation (1):

$$\text{(1)} \quad \mathrm{topic-sim}(\mathrm{tw_i}, \mathrm{P_{u_i}}) = \frac{\mathrm{topic}_{\mathrm{tw_i}} . \mathrm{topic}_{\mathrm{P_{u_i}}}}{||\mathrm{topic}_{\mathrm{tw_i}}||.||\mathrm{topic}_{\mathrm{P_{u_i}}}||}$$

Using the LDA method, the topic of each tweet can be obtained in the form of a topic containing a group of words that are likely to occur in this topic. As a result, using this method, the topic of the tweet is found and the category of words related to this topic is determined. If the tweet $tw_i$ is first deleted by a standard method of resident words and the set $tw_i = \{w_1, w_2, \ldots, w_m\}$is a set of words used in the tweet $tw_i$, and the set $M_{topic\#} = \{w_1, w_2, \ldots, w_L\}$is a set of words related to the topic $tw_i$, the following relationship determines the novelty of the tweet:

$$\text{(2)} \quad \mathrm{Novelty} = \frac{\mathrm{topic-sim}(\mathrm{tw_i}, \mathrm{P_{u_i}})}{\mathrm{sim}(\mathrm{M_{topic\#}} - \{\mathrm{M_{topic\#}} \cap \mathrm{tw_i}\}, \mathrm{P_{u_i}})}$$

In equation (2), $M_{topic\#} - \{M_{topic\#} \cap tw_i\}$is a set of related words that is not in the tweet. As can be seen in Equation (2), the more similar the topic-related words are to the topics discussed by the user, the less likely it is that the tweet will be novel. The novelty of the tweet is by definition directly related to the user-created tweets.

### 3.1.2 Checking the location of a tweet

In addition to the novelty of the tweet, the relation of a tweet with the user can be a key point in the content value of this tweet by the user. In this paper, in addition to calculating the lexical similarity between tweet $\boldsymbol{tw_i}$ and the tweets used by the user, a geographical similarity is considered. Typically, tweets that address regional issues are more effective than tweets from other regions. Therefore, in this paper, maxmind data set including 4 million names of cities and different regions along with additional information about the country, location, etc. is used to find words related to cities and geographical regions. To find words related to cities from the tweet $\boldsymbol{tw_i}$ , all the words in the tweet are used except for the extra words (even hashtags with the # symbol removed). If $\boldsymbol{LOC_{tw_i}}$ is a set of cities plus the country and region used in the tweet and $\boldsymbol{LOC_{U_i}}$ is a set of cities, countries and regions used by the user $\boldsymbol{U_i}$ in all created tweets, the relationship (3) indicates the geographical relationship between the tweet $\boldsymbol{tw_i}$ and the user $\boldsymbol{U_i}$:

$$\text{(3)} \quad \underset{\mathrm{u_i \in U \& tw_i \in P}}{|(\mathrm{Loc_{tw_i}} \cap \mathrm{Loc_{u_i}})|}$$

In (Gou et al., 2011) it is shown that logical news and information can be disseminated more than positive events. Therefore, in order to analyze the content of a tweet to measure the content value, in this paper, the negativity of the tweet is determined. In the method used to measure the negative value of the tweet, an attempt was made to use a simple method for online execution. The University of Michigan Twitter dataset was used for this purpose. In this data set, 1,578,627 labeled tweets (label 1 for positive tweets and 0 for negative tweets) were collected. Using the Python NLTK natural language processing tool kit and the naïve bayes method, sentiment analysis was conducted. To obtain the separator, 1/10 of the data was used for testing and the rest for training. As a preprocessing, two steps of tokenization and normalization are performed on the data set. Emoticons and abbreviations (OMG, ASAP, WTF…) are considered as separate tokens in the tokenization step. In the normalization phase, uppercase words are converted to lowercase letters (for example I LOVE it !!!!) and the repetition of letters in tweets (for example I am happyyyy !!) is removed.

### 3.1.3 Checking the growth of a tweet

Another parameter that determines the content value of a tweet is the growth of the tweet's topic. Tweet topics taken from posts outside the microblog environment can have a direct impact on the retweeting behavior of the tweet. In this work, using the LDA method mentioned above, topics related to a tweet are extracted. The growth rate of the tweet topic over a period of time (24 hours) is calculated using Equation (4):

$$\boldsymbol{Topic_{growth}} = \left[\frac{\mathbf{1}}{\mathbf{24}}\left(\frac{\sum_{w_i \in M_{topic\#}} \#\boldsymbol{tW}^{t}_{w_i} - \sum_{w_i \in M_{topic\#}} \#\boldsymbol{tw}^{t-24}_{w_i}}{\sum_{w_i \in M_{topic\#}} \#\boldsymbol{tw}^{t-24}_{w_i}}\right)\right] \times \mathbf{100} \quad (4)$$

In (4), $\boldsymbol{M_{topic\#}}$ is a set of words related to the topic of the tweet $\boldsymbol{tw_i}$ and # $\boldsymbol{tw^{t-24}_{w_i}}$ is number of tweets from the last 24 hours that contain the words $\boldsymbol{tw_i \epsilon M_{topic\#}}$ for all tweets $\forall \boldsymbol{tw_i \epsilon p}$ .

### 3.1.4 Tweeter credibility

The tweet feature is also used to complete the content value of the tweet. This feature was selected to complement the content in such a way that it can be used online and efficiently. Many features have been used to predict retweeting behavior in numerous articles, however , in selecting these features, two aspects of operational capability and relevance to content value were considered.

According to the retweeting statistics of a post, a feature is introduced which in this article is called the weight of a tweet. The weight of the tweet is directly proportional to the credibility of the tweeter and the users who retweeted the tweet. If $U^{RT}_{tw_i} = \{u_1, u_2, \dots, u_R\}$ is the set of users who retweeted the tweet $\boldsymbol{tw_i}$ by time t, equation (5) determines the weight of the tweet:

$$\boldsymbol{weight(tw_i, t)} = \left[\sum_{u_j \in U^{RT}_{tw_i}} \boldsymbol{log}\left(\frac{\boldsymbol{\#f(u_j)}}{\boldsymbol{\#E(u_j)}}\right) + \boldsymbol{log}\left(\frac{\boldsymbol{\#f(Au_{tw_i})}}{\boldsymbol{\#E(Au_{tw_i})}}\right)\right] \times (\boldsymbol{t} - \boldsymbol{t_{tw_i}})^{-\boldsymbol{\beta}} \quad (5)$$

$\#f(u_j)$ and $\#E(u_j)$ are the number followers of the user $\boldsymbol{U_j}$ and the number of followings by the user $\boldsymbol{U_j}$, respectively. t indicates the calculation time and $\boldsymbol{Au_{tw_i}}$ indicates the tweeter of tweet $\boldsymbol{tw_i}$. Due to the mutual following relationship, it is not possible to determine the user or tweeter credibility only based on #. The higher the number of followers of user $\boldsymbol{U_j}$ than the number of followings by user $\boldsymbol{U_j}$, the greater the credibility of user $\boldsymbol{U_j}$.

One of the most important and decisive aspects for the content value is the age of the tweet. According to the evaluation data set of this article, the rate of retweeting a tweet is very similar to the power point in terms of its age. β is the power of the time elapsed since the creation of the tweet $tw_i$, which is applied to the weight of the tweet according to the power point. Since a tweet is created, the weight and content value of that tweet decreases over time.

### 3.2 Effective learning using a deep learning-based method

In this paper, using different architectures, a deep learning model is presented to develop a predictive system for predicting users' retweeting behavior. In the training phase, a deep auto-encoder algorithm is trained using normal network observations to generate the initial parameters. The basic parameters include weight and bias, and this algorithm trains a deep representation of normal observations. These parameters are used as an initial step in training a standard deep feed forward neural network (DFFNN) to learn existing feature coefficients and predict future user behaviors in facing a post. In the experimental step, the standard deep feed forward neural network is used to learn the coefficients and form a classification. The various hidden nodes in this technique learn the representation of deep features professionally and receive the most important features by transforming the high-dimensional data to the low-dimensional ones based on the reduction of the hidden layer. The details of the proposed method are described in detail in the following.

#### 3.2.1 Deep Feed Forward Neural Network (DFFNN)

Typically, the deep feed forward neural network is considered as an intelligent neural network that has one input layer, more than one hidden layer, and one output layer with direct connections without cycles between them. Each hidden layer of nodes displays the abstract features based on the output of the previous level, which are automatically determined and collected in several layers to produce the output. To train this method, a back propagation algorithm with stochastic gradient descent is used. In this deep learning algorithm, the input data is placed inside the input layer and then propagated to the hidden layer, the output of which is a non-linear transfer of data and passed to the output layer. The loss or error propagation function, which is the difference between the predicted output and the actual output, is used to evaluate the performance of the model and its value is propagated as a back propagation through hidden layers to update the weights. The loss function or cost function actually displays the amount of error each time the neural network is executed for the training data. The loss function is calculated based on a data sample or a small set of training data instead of the total data, with the weights updated after each sample processed to fit the model. The supervised learning process in this algorithm depends on the stochasticity of the initial neural network parameters, which tends to place the model in a minimal local solution with poor settings. To better converge and improve supervised learning results, unsupervised pre-learning techniques, especially auto-encoder networks, can be used to generate initial parameters.

#### 3.2.2 Deep Auto-Encoder (DAE)

DAE is a feed forward neural network algorithm for effective coding learning using an unsupervised technique. This algorithm represents a data set (x) through the approximate learning of a discriminant function, where the output (xˆ) is the same as the input (x), ie $x \rightarrow x\hat{}$. Its schematic structure consists of vectors (x(i)) in the input layer and more than one hidden layer with a nonlinear activation function. Hidden layers are used to learn a concise representation of input data by neurons smaller than the input layer. As a result, it learns the most important features, reduces dimensions, and displays an abstract of the input data. Finally, the output layer (xˆ (i)) is displayed as an approximation of the input. The simplest architecture of an auto-encoder consists of an input layer, a hidden layer and an output layer. It is assumed that the learning data (x (i)) has n samples, so that each x (i) (i∈ (1,…., N) has many dimensions and there is a one-dimensional eigenvector (d0) and the activation function Tanh is used and calculated as Equation (6):

$$T(t)=\frac{1-e^{-2t}}{1+e^{-2t}} \tag{6}$$

The auto-encoder algorithm has two main parts: encoder and decoder. To map the input vector (x (i)) to a hidden layer represented by (z (i)), a definite map called the encoding process (fθ) is used, and the dimensions (x (i)) are reduced for obtaining the correct number of codes as Equation (7):

$$f_\theta(x^{(i)}) = T\ (Wx^{(i)} + b) \tag{7}$$

Where W is the weight matrix of d0 × dh, dh is the number of neurons in the hidden layer (dh <d0), b is the bias vector, T is the Tanh activation function and θ are parameters of mapping [W, b]. To reconstruct the input as an approximation (xˆ (i)), the result of the hidden layer display is mapped and the decoding process with definite mapping (gθ´) is calculated as Equation (8):

$$g_{\theta'}(x^{(i)}) = T(w'_z{}^{(i)} + b) \tag{8}$$

Where w is the weight matrix dh × d0, b is the bias vector and θ are all parameters of mapping [W, b].

The input is formed in a concise display to proportional to the hidden layer, and then the data used as input is employed to reconstruct the original data. The learning process minimizes the reconstruction error (for example, the distinction between the original data and its small-scale reconstruction) and is calculated for a sample or small set s in the form of equations 9 and 10:

$$E\ (x, x\hat{}) = ½ \sum_i^s ||\ \boldsymbol{x(i)}\ - \boldsymbol{x\hat{}(i)}\ ||^2 \tag{9}$$

$$\theta=\{w,b\}=\text{argmin}\theta\ E\ (x, x\hat{}) \tag{10}$$

The DAE architecture consists of three hidden layers, an encoder, a bottleneck (which contains fewer nodes than the previous layers and is used to display input data with reduced nonlinear dimensions, in which the number of nodes indicates the number of dimensions) and a decoder. This model is potentially utilized using nonlinear principal component analysis (PCA) to reduce dimensions. In order to predict the retweeting behavior of users on social networks, we propose a prediction method that includes learning and testing phases and which is described below.

## 3. 2.3 Phases of learning and testing of retweeting behavior prediction system based on deep learning

The Deep Feed Forward Neural Network (DFFNN) and the Deep Auto-Encoder (DAE) discussed in the previous sections are the basic mechanisms for predicting retweeting behavior of users based on deep learning. The structure of the DAE-DFFFN network is shown in Figure1.

In the learning phase, the normal learning dataset without label A, the learning dataset with label B, in which ACB is a deep auto-encoder (DAE) with a single-layer bottleneck containing only normal records (A) without any unusual vectors for learning, is trained and it covers the most important features for displaying normal patterns. This training with all the data, where the input to the network (x (i)) passes through three hidden layers, including a bottleneck for reconstruction (xˆ (i)), is done where We, Wn and Wf are the weights of deep feed forward neural network (DFFNN), the deep auto-encoder (DAE) and Final Prediction Model described in Figure 1.

In the encoding step, the input layer is processed in the first hidden layer using equations (6) and (7). In the bottleneck step, a low-dimensional nonlinear transformation of the input feature is performed to extract features effective in users' retweeting decision. Then, in the decoding step, the last hidden layer in the bottleneck feature is used to approximate the input repetition using Equations (6) and (8). Then the back propagation of stochastic gradient descent is used to reduce the loss function or the cost function, ie the mean square error between (x (i)) and (xˆ (i)) using equations (9) and (10). A key process for unsupervised learning of predicting user retweeting behavior based on the proposed deep learning is presented in Figure 2:

Then the training model for the starting point and initialization of weight and bias parameters for learning for the deep network of supervised learning and setting network model using labeled learning datasets (B ((x (i)), (y (i)) ))) is used. The same steps as before are followed to learn and validate the method of predicting user retweeting behavior based on the proposed deep learning on data set B, which includes data on retweeting and non-retweeting behaviors to test the accuracy of the proposed method. In more detail, a network model based on stochastic gradient descent back propagation mechanism to minimize the loss function with the mean squared error calculated from the difference between the target output value (y (i)) and the predicted output gθ´ (x (i)) is trained is where gθ is a hypothetical function that results from an estimated output.

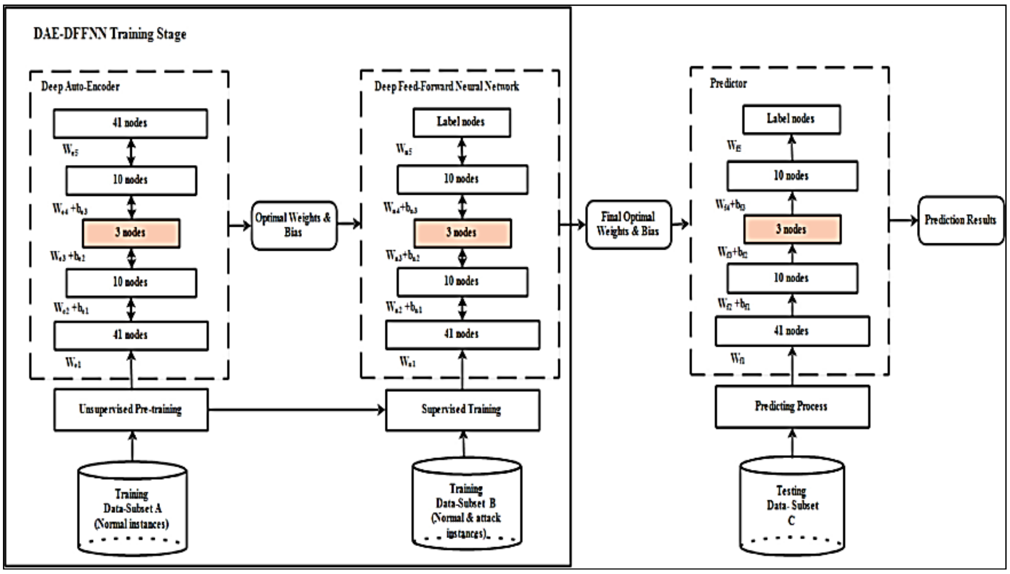


*Fig. 1: The structure of the DAE-DFFFN network i*

```
Input: training dataset (A) with n umber of samples (n) of
  (x^(i)), where i ∈ (1, …., n).
Output: parameters θ = {W, b}
Begin
  Initialize {W, b};
    repeat
      For each record (x^(i)), do
        compute the activation ( z^(i)) in/at hidden layer and
        give output (x̂^(i)) to outhe tput layer.
        compute the training error (E(x^(i), x̂^(i)).
        Back-propagate E and update parameters
        θ = {W, b};
    End
until converged
end
```

***Fig. 2: Unsupervised learning phase for predicting republishing behavior of users based on deep learning***

In the test phase, after the parameters are automatically trained in the learning phase, the sample of the new data set ($\boldsymbol{C} \subset_{\neq} [\boldsymbol{A}, \boldsymbol{B}]$) based on the final constructed network model is tested. Each input record (xˆ (i)) is matched to the input layer with the initial weight and bias value (θf = Wf, bf) and then the input data is processed through the hidden layers. Finally, the output layer predicts the input data class as retweet or non-retweet based on the estimated value of the loss function of each class.

### 3.3 Proposed framework for the system

This paper proposes an efficient retweeting behavior prediction method for studying users' behaviors in facing various tweets, as shown in Figure 3. This method is called DLMD.As can be seen in the figure below, the architecture includes the steps of learning and testing. Data preprocessing, which includes feature transformation and normalization, is the first step in the proposed deep learning-based retweeting behavior prediction mechanism that examines and selects important information from large-scale data in a social networking environment.

Feature transformation: Since the proposed method accepts only numerical features, each value of the symbolic feature changes to a numeric value. For example, in the collected data set, there are a number of symbolic features such as the tweet negativity, including negative, positive, or neutral tweets, which are mapped to values 1,2, and 3.

Feature normalization: Because deep learning depends on weight, different feature scales can bias data into specific layers, causing certain weights to be updated faster than other weights. As a result, this needs to be corrected by statistical normalization so that the z-score for each feature value (v (i)) is calculated as Equation (11):

$$\boldsymbol{Z}^{(i)} = \frac{\boldsymbol{v}(\boldsymbol{i}) - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \quad (11)$$

Here μ is the mean value of n for the given feature v (i) (i∈1, 2,3,4, ..)) and σ is the standard deviation. Since the network data contains a space with high dimensions, it is necessary to reduce data

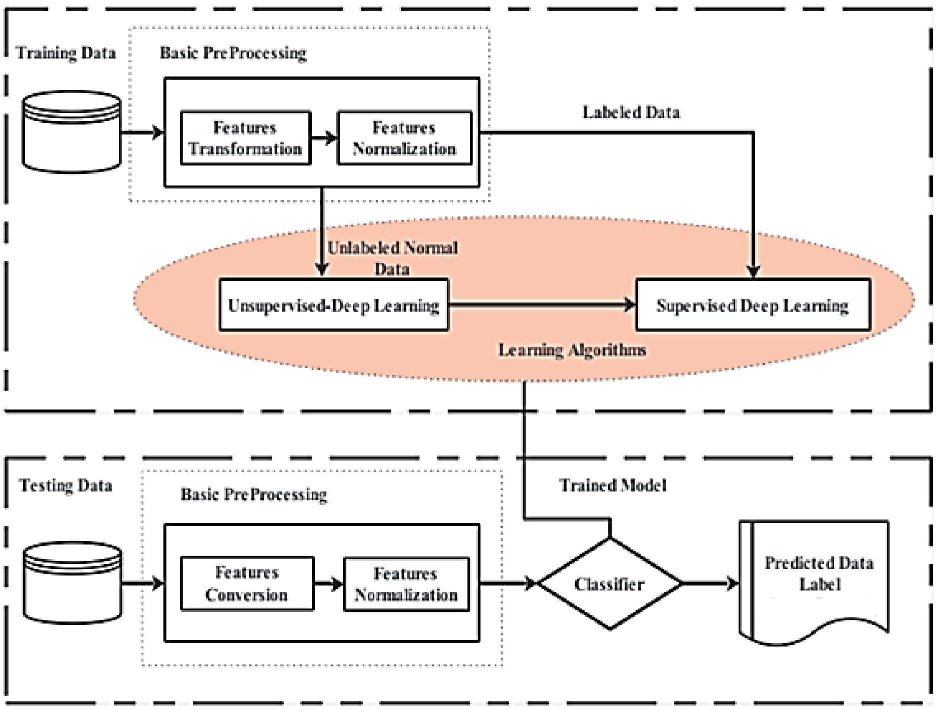


*Fig. 3: The structure of proposed method(DLMD)*

dimension for improving computational resources in order to design a method for predicting user retweeting behavior on a lightweight scalable basis. In more detail, there is a nonlinear function in the model that encodes a large number of features into a lower set of features in the reduced hidden layer, so the feature reduction is done without the need for human knowledge. The purpose of reducing the DAE-DFFFN feature is to limit the ambiguous structure in the input distribution and finding a well-designed display with respect to high-level learning, importance filter and reduced feature. Therefore, an unsupervised learning process is used to observe users' retweeting behavior for specifying initial estimation of weight and bias parameters of an input given by the standard DFFNN in order to reduce the processing time of this model. The parameters are also re-adjusted using labeled data (normal and destructive) in the supervised deep learning method, with an evaluated final learning model based on new data samples obtained during the test phase.

## 4- Evaluation

In this Section , we explain in details how the datasets were collected and how the experiment were conducted. We collected four different datasets from twitter and performed different experiments on them to see to what extent we can predict the popularity of tweets in the Twitter social network and maximize diffusion.

### 4-1 Dataset

Twitter is an information exchange network that produces 200 million tweets per day. To be able to test our proposed method, we created four different datasets using the twitter streaming API. Having four different datasets allows us to test our method on different situations to see how well our method can be generalized. These datasets are illustrated in Table 1.

*Table 1: the detail of dataset*

| Dataset | Description | # Tweets | # User | # Retweet |
|---|---|---|---|---|
| Covid 19 | Corona virus out break in China | 24399 | 345709 | 5600553 |
| George Floyd | George Floyd, a 46-year-old black man, was murdered in Minneapolis, | 25679 | 60267 | 851071 |

| | | | | |
|---|---|---|---|---|
| Australia's worst natural disaster | Nearly 186,000 sq km area was burnt, and over a billion wild | 55240 | 56489 | 158133 |
| 😊 | All tweets contain :) | 13956 | 1162 | 21578 |

## 4-2 Evaluation Metrics

We use Specificity, accuracy and F1-measures to evaluate the performance of the method. The Specificity, accuracy and f-measure are calculated according to equation 12 , 13 ,14 ,15and 16:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{12}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{13}$$

$$\text{Precison} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{15}$$

$$\text{Fmeasure} = \frac{2 * (\text{Precison} * \text{recall})}{\text{Precison} + \text{recall}} \tag{16}$$

In equation 12 , 13 ,14 ,15and 16 variables are showed by table 2.

*Table 2: The variable of recall and precision*

| | Prediction Condition | | |
|---|---|---|---|
| True Condition | Condition Positive | TRUE Positive | False Negative |
| | Condition Negative | False Positive | TRUE Negative |

## 4-2 The Result of Evaluation

In this section, the proposed method is compared with other methods based on Specificity, Accuracy and F-measure. Given is a table (3) illustrating the proposed method is compared with the other four methods in terms of Specificity. As shown in table (3), the DLMD performed best. In terms of Specificity, logistic regression has the best performance after DLMD on average. In this section, the support vector machine has the worst performance.

As can be seen in table (4) ,the DLMD method has performed best in accuracy, also the second most great performance is SVM while the worst performance is Decision Tree.

According to table (5), the F-measure of different methods are compared. The F-measure is weighted average of precision and recall. Proposed method has best performance while decision tree has worst performance. Although logistic regression has low accuracy its F-measure has high value.

## 5- Conclusion

Diffusion prediction on the social networks is one of those popularity studies that many papers focus on. Recent studies show that there are many methods for diffusion prediction based on feature, timeseries etc. In this paper, we proposed a method based on deep learning on content value of a tweet and it is called DLMD.

We experimentally test our proposed method with four datasets that they collected twitter steaming API. The proposed method compared with other methods based on Specificity, accuracy and F- Measure. The performance of proposed method is better than other methods.

One of the best areas for future work is to "provide a mechanism for selecting effective multimedia content in online communities." Today, video sharing portals such as YouTube has millions of daily visitors. The huge financial cycle of these portals has been made possible through advertising. Increasing the number of videos viewed on these systems has increased revenue. According to research, the effectiveness of multimedia content is much higher than textual content, which makes the use of this type of content for advertising more appropriate. Today, with advances, features such as note-taking, tagging, liking, sharing, and more on videos have emerged. This information can be very useful for selecting effective multimedia content. As a suggestion, you can create a model of users by measuring the viewing time, reviewing the

*Table 3: Comparison of the Specificity criteria of the proposed method with other methods*

| Specificity Measure | Average | First Dataset | Second Dataset | Third Dataset | Fourth Dataset |
|---|---|---|---|---|---|
| DLMD | 0.709166068 | 0.5707317 | 0.869600 | 0.74306657 | 0.6532660 |
| Logistic Regression | 0.620629173 | 0.789980 | 0.33345689 | 0.4900898 | 0.86899000 |
| Decision Tree | 0.55808887 | 0.5563305 | 0.4476707 | 0.3245765 | 0.90377778 |
| SVM | 0.529732155 | 0.6456798 | 0.23447808 | 0.47328085 | 0.76548989 |

*Table 4: Comparison of the accuracy criteria of the proposed method with other methods*

| accuracy Measure | Average | First Dataset | Second Dataset | Third Dataset | Fourth Dataset |
|---|---|---|---|---|---|
| DLMD | 0.861216068 | 0.7707317 | 0.877800 | 0.94306657 | 0.8532660 |
| Logistic Regression | 0.55562005 | 0.789980 | 0.4334500 | 0.4300567 | 0.5689935 |
| Decision Tree | 0.400144075 | 0.456890 | 0.3856009 | 0.357098 | 0.4009874 |
| SVM | 0.673276 | 0.6706798 | 0.577848 | 0.679087 | 0.76548989 |

*Table 5: Comparison of the F-measure criteria of the proposed method with other methods*

| F- Measure | Average | First Dataset | Second Dataset | Third Dataset | Fourth Dataset |
|---|---|---|---|---|---|
| DLMD | 3.25 | 3.6 | 3.12 | 3.20 | 3.11 |
| Logistic Regression | 2.39 | 2.4 | 2.18 | 2.35 | 2.65 |
| Decision Tree | 0.68 | 0.71 | 0.65 | 0.76 | 0.60 |
| SVM | 1.93 | 1.98 | 1.89 | 1.56 | 2.30 |

subject, hashtags and notes of a video. Using this modeling, user feedback in the face of other videos is predictable. This helps increase the revenue and revenue of video sharing systems and increase user satisfaction.

## Reference

Alshaabi, T., Dewhurst, D. R., Minot, J. R., Arnold, M. V., Adams, J. L., Danforth, C. M., & Dodds, P. S. (2020). The growing echo chamber of social media: Measuring temporal and social contagion dynamics for over 150 languages on twitter for 2009–2020. arXiv preprint arXiv:2003.03667.

Anderson, R. M., & May, R. M. (1992). Infectious diseases of humans: dynamics and control. Oxford university press.

Bhowmick, A. K., Gueuning, M., Delvenne, J. C., Lambiotte, R., & Mitra, B. (2019). Temporal sequence of retweets help to detect influential nodes in social networks. IEEE Transactions on Computational Social Systems, 6(3), 441-455.

Cannarella, J., & Spechler, J. A. (2014). Epidemiological modeling of online social network dynamics. arXiv preprint arXiv:1401.4208.

Cao, Q., Shen, H., Cen, K., Ouyang, W., & Cheng, X. (2017, November). Deephawkes: Bridging the gap between prediction and understanding of information cascades. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 1149-1158).

Cerchiello, P., Giudici, P., & Nicola, G. (2017). Twitter data models for bank risk contagion. Neurocomputing, 264, 50-56.

Chen, W., Wang, C., & Wang, Y. (2010, July). Scalable influence maximization for prevalent viral marketing in large-scale social networks. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1029-1038).

Fang, A. (2021). The influence of communication structure on opinion dynamics in social networks with multiple true states. Applied Mathematics and Computation, 406, 126262.

Feng, L., Hu, Y., Li, B., Stanley, H. E., Havlin, S., & Braunstein, L. A. (2015). Competing for attention in social media under information overload conditions. PloS one, 10(7), e0126090.

Gou, L., Zhang, X., Chen, H. H., Kim, J. H., & Giles, C. L. (2010, June). Social network document ranking. In Proceedings of the 10th annual joint conference on Digital libraries (pp. 313-322).

Hansen, L. K., Arvidsson, A., Nielsen, F. Å., Colleoni, E., & Etter, M. (2011). Good friends, bad news-affect and virality in twitter. In Future information technology (pp. 34-43). Springer, Berlin, Heidelberg.

Hong, L., Dan, O., & Davison, B. D. (2011, March). Predicting popular messages in twitter. In Proceedings of the 20th international conference companion on World wide web (pp. 57-58).

Kempe, D., Kleinberg, J., & Tardos, É. (2003, August). Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 137-146).

Leskovec, J., Adamic, L. A., & Huberman, B. A. (2007). The dynamics of viral marketing. ACM Transactions on the Web (TWEB), 1(1), 5-es..

Li, C. T., Shan, M. K., Jheng, S. H., & Chou, K. C. (2016). Exploiting concept drift to predict popularity of social multimedia in microblogs. Information Sciences, 339, 310-331.

Li, C., Ma, J., Guo, X., & Mei, Q. (2017, April). Deepcas: An end-to-end predictor of information cascades. In Proceedings of the 26th international conference on World Wide Web (pp. 577-586).

Li, C., Ma, J., Guo, X., & Mei, Q. (2017, April). Deepcas: An end-to-end predictor of information cascades. In Proceedings of the 26th international conference on World Wide Web (pp. 577-586).

Li, M., Wang, X., Gao, K., & Zhang, S. (2017). A survey on information diffusion in online social networks: Models and methods. Information, 8(4), 118.

Liu, C., & Zhang, Z. K. (2014). Information spreading on dynamic social networks. Communications in Nonlinear Science and Numerical Simulation, 19(4), 896-904.

Liu, D., Yan, E. W., & Song, M. (2014). Microblog information diffusion: Simulation based on sir model. J. Beijing Univ. Posts Telecommun, 16, 28-33.

Mishra, S., Rizoiu, M. A., & Xie, L. (2016, October). Feature driven and point process approaches for popularity prediction. In Proceedings of the 25th ACM international on conference on information and knowledge management (pp. 1069-1078).

Mishra, S., Rizoiu, M. A., & Xie, L. (2016, October). Feature driven and point process approaches for popularity prediction. In Proceedings of the 25th ACM international on conference on information and knowledge management (pp. 1069-1078).

Ni, L., Chen, Y. W., & de Brujin, O. (2021). Towards understanding socially influenced vaccination decision making: An integrated model of multiple criteria belief modelling and social network analysis. European Journal of Operational Research, 293(1), 276-289.

Sandberg, J., Park, C., Rytina, S., Delaunay, V., Douillot, L., Boujija, Y., ... & Senghor, A. (2019). Social learning, influence, and ethnomedicine: Individual, neighborhood and social network influences on attachment to an ethnomedical cultural model in rural Senegal. Social Science & Medicine, 226, 87-95.

Stowe, K., Palmer, M., Anderson, J., Kogan, M., Palen, L., Anderson, K. M., ... & Lazrus, H. (2018, August). Developing and evaluating annotation procedures for twitter data during hazard events. In Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018) (pp. 133-143).

Trzciński, T., Andruszkiewicz, P., Bocheński, T., & Rokita, P. (2017, June). Recurrent neural networks for online video popularity prediction. In International Symposium on Methodologies for Intelligent Systems (pp. 146-153). Springer, Cham.

Wang, C., Yang, X. Y., Xu, K., & Ma, J. F. (2014). SEIR-based model for the information spreading over SNS. Acta Electronica Sinica, 11, 2325-2330.

Vallet, D., Berkovsky, S., Ardon, S., Mahanti, A., & Kafaar, M. A. (2015, October). Characterizing and predicting viral-and-popular video content. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (pp. 1591-1600).

Wang, Q., Lin, Z., Jin, Y., Cheng, S., & Yang, T. (2015). ESIS: emotion-based spreader–ignorant–stifler model for information diffusion. Knowledge-based systems, 81, 46-55.

Wang, S., Cuomo, S., Mei, G., Cheng, W., & Xu, N. (2019). Efficient method for identifying influential vertices in dynamic networks using the strategy of local detection and updating. Future Generation Computer Systems, 91, 10-24.

Wu, S., Tan, C., Kleinberg, J., & Macy, M. (2011, July). Does bad news go away faster?. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 5, No. 1).

Xu et al. (Li etal, 2017) found that information diffusion is not only related to a user's behavior S-SEIR model is proposed .it focused on the value of the information.

Yang, J., & Counts, S. (2010, May). Predicting the speed, scale, and range of information diffusion in twitter. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 4, No. 1).

**HOW TO CITE THIS ARTICLE:**
*Safearyan H., Tarokh M., Afshar Kazemi M. (2022). Prediction of Message Diffusion: A Deep Learning Approach on Social Networks, 3(4): 33-46.*

**DOI:**

**Url:** **https://ijfaes.srbiau.ac.ir/article_16607.html**
**Journal homepage:** **https://ijfaes.srbiau.ac.ir**