

# A New Approach in Bloggers Classification with Hybrid of K-Nearest Neighbor and Artificial Neural Network Algorithms

Farhad Soleimanian Gharehchopogh<sup>1\*</sup>, Seyyed Reza Khaze<sup>2</sup> and Isa Maleki<sup>3</sup>

<sup>1</sup>Department of Computer Engineering, Hacettepe University, Beytepe, Ankara, Turkey; bonab.farhad@gmail.com

<sup>2</sup>Young Researchers and Elite Club, Dehdasht Branch, Islamic Azad University, Dehdasht, Iran; khaze.reza@gmail.com

<sup>3</sup>Young Researchers and Elite Club, Urmia Branch, Islamic Azad University, Urmia, Iran; maleki.misa@gmail.com

## Abstract

Blogs are one of the effective tools of web2 which are considered as one of the major module and of social and interactive capabilities in making IT world wonderful for the cyber and virtual living. Two methods were used in this paper: K-Nearest Neighbor (KNN) and Artificial Neural Networks (ANNs). These methods are classified based on Kohkiloye and Boyer Ahmad province bloggers dataset considering input features of each blogger to the other methods and previously provided algorithms as more optimal. Our simulation and experiments not only provide hopeful results but also higher anticipation and classification rate.

**Keywords:** Artificial Neural Networks, Bloggers Classification, Decision Tree, K-Nearest Neighbor

## 1. Introduction

The wonders of IT advancements change the people's perception of the living universe and invite them to the modern and attractive digital world<sup>1</sup>. Nowadays, web2 has been spread in all levels of IT and cyber space world. Without learning technical education, the users can share their information, diaries and daily issues and even thoughts by using blogs which is one of the most popular tools of web2 to provide contexts in virtual space<sup>2</sup>. In blogs, the sent articles of the users are displayed as time-reverse. It means that the new articles are available at the top. Moreover, the user can links to other web sites and blogs, too<sup>3</sup>. In general, the blogs are administrated by one person but it can be managed by a group of writers. Due to the broad capabilities of the blogs, they are like online diaries which easily accessible and free of charge<sup>4</sup>. The users can get proper feedback and related ideas about their sent articles or attend in social and political discusses with the others through their blogs, so in compared with web2 tools, blogs have better social aspects and capabilities

which provide an opportunity for the bloggers to follow their various social goals<sup>2,5</sup>. Virtual discuss and communication among users through blogs indicate a virtual interaction and cooperation which resulted in increasing popularity of blogs all over the world. Due to the broad studies of the research institution, it can be seen millions of blogs in internet<sup>6,7</sup>. It is obvious that there is a special relation between blogs and political, social and cultural patterns of the countries so; bloggers classification is also resulted in identifying community individuals and their ideas, thoughts and approaches appropriate to their living area. By the way, we can get familiar with public aspects and different people ideas and thoughts<sup>1</sup>. So, understanding that who will tend to be professional blogger and what is their goal can help us to have better understanding of countries circumstances and get familiar with their people and communities. We review the subject literature and performed researches about blog and blogging in this paper. In following, KNN algorithm is described and the classification method based on the blogging trend is illustrated and attempted to improve the classification results based

\*Author for correspondence

on C4.5 decision trees algorithm<sup>1</sup>. Kohkiloye and Boyer Ahmad Province bloggers' dataset<sup>1</sup> was used as our experiment population and also artificial neural network method for classification. Moreover, we attempted to improve the accuracy of classification as well as the gained results of the nearest neighbor algorithm in this paper. In the final section, the gained results of two methods classification are presented in compared with the previous studies.

## 2. Previous Work

In <sup>1</sup>pointed out the blogs as dynamic bridge among the internet users; the blog capabilities as a free public media and the highway of communication and idea exchange. By providing a method based on decision making tree and using C4.5 algorithm, an attempt was made to classify and anticipate blogging trend factors. The decision making tree makes a model before anticipation which can be seen its trend factors and potential relations capabilities. The precision rate of the bloggers in this respect was 82%.

In <sup>8</sup>classified political blogs and their ranking by implementing their innovational algorithm. The provided algorithm was a quasi-supervision one and classification precision of the two used dataset was 81.9 and 84.6, respectively. The effective parameters on their experiments were the numbers of blogs graph nodes in each group, metric development, inclusion and deletion of resolution phase. By using multi rank algorithm in all phases, the writers choose groups which have higher page rank. The advantage of this method is that the classification errors have been made for the lower ranked blogs. It can be used for classifying political blogs which have related structure and high classification rate. The other researchers <sup>9</sup>point out to the blogs as online diaries which are updated interactively and regularly. The writers emphasize that the websites are formed based on subjects and the need for methods to data recovery of potential knowledge in online web resources. They consider article classification as the first step. By using linguistic knowledge, classification of articles posted on blogs is an improvement. Regardless of the blogs length and writing style, Jung et.al<sup>10</sup> have attempted to identify blog mode which included a general public knowledge and emotional norms expressions. In this method, the unique features of the blog are taken and a simple statistics are saved. This statistics include words and other data which is useful for SVM classification algorithm. It is attempted to perform paragraph division based on mood flow analyses of blogs according to a modified

version of Guess Mood as well as blogs mood identifying. In order to assess mood in this article, the formed texts members of real blogs are made as semi-automatic and finally 4 moods are taken. These 4 items include happy, sad, fear and anger moods. Other researchers <sup>11</sup>have studied Cool Blogs classification by using positive and non-labeled samples. By providing innovative algorithm, they have complemented blogs classification. It is used non-labeled data as well as positive samples to predict whether a blog includes in Cool Blog classification in this algorithm. The provided algorithm has been used weighting techniques to assign weight for each perceived labeled samples which have negative effect in educational set. They also pointed out any weak classification which must be assessed as virtual function scale based on total labeled sample or the best repeated efficiency scale. To identify the accuracy of estimation in classified classes, the experiment sample used the function of majority voting in all classes. The provided algorithm could accurately identify blog potential class which was not comparable with traditional learning methods. It is necessary to note that this algorithm gives better results than the others to classify Cool Blog. In another article<sup>12</sup>, blogs polarization have been widely reviewed and that a detailed look in polarization prospect about the particular subject could be taken from a series of blogs. A cascade set of independent texts classification was placed above text recovery engine with limited effects. In this feature, they begin to review more precise and vague aspects identifying of polar texts classification beyond the simple one. They also come to investigate shared analyses of subjective condition and polar texts. In <sup>13</sup>reported the style difference results in blogging based on age and gender. It is based on 2 independent and interactive features. The first feature is to use slang words which they use as a new concept to study blogging style. For the second feature, they analyze random variety of sentences which is used among different age groups. They emphasize that the latter feature increase the precise prediction of the age and gender of the bloggers. In their methods, gender prediction had higher precision compared with identifying groups' age. In <sup>14</sup>suggested a new method of blog interests which is based on Chinese blog text classification. Before text classification, they started to pre-process of texts to display. Then, the texts are displayed in space vector and began to classify them by using support vector machine. It filtered unrelated and redundant texts and then classified based on interests. The results showed that support vector machine

is an effective algorithm to identify interests and data of blogs. In <sup>15</sup>based on classification method, provided the state of feelings in linked blogs. At the first step, they began to find related blogs in blog space. As the blogs are generally linked via many other blogs which trust them, they noted that there are passages in blogs that the writer proposed to state his/her ideas and feelings about links and related blogs. It has been done by providing a method in the entry of Japanese blogs to extract these passages automatically and then began to classify links by using available data in extracted area. It has been used estimation precise method and recall of base method for effectiveness. In <sup>16</sup>emphasized that blogs are the produced contents of the internet users and allocate to discussions about different subjects. The writers emphasized on knowledge-oriented Wikipedia which has been used widely by researchers to adopt issues such as feature selection and classification. Their empirical results have developed the percent of classification in this environment. The other research group<sup>17</sup> also noted to the blogs' importance and popularity in recent years and considered them as a valuable data resource. By providing 2 stages of new and modern framework, they begin to find meaningful phrases of blog posts. This system has subjected the blogs classification and performed it by subjecting in text-dependent topics. Another researcher <sup>18</sup>considered blog analyses as an important area of informatics behavior. The writer considers the importance of this topic as far as he/she believes that the blogs analyses will result in anthropology of social media. In <sup>19</sup>applied gender approach and classification by using Naive Bayes classification algorithm. The authors also used features such as background color and small images beside the noted method to classify. The other researchers' findings in <sup>20</sup>pointed out the exponential growth of blogs; most directories seem to help users to classify blogs as subjective. By emphasizing on labels which are usually used to describe blogs, they begin to survey label classification effect in blogs classification. For better description, they Label expansion algorithm development. The fact which can be taken from other research group<sup>21</sup> is that the difficulties of blog classification are due to the inequality of writing style. The writers who noted to it emphasize that by increasing growth of blogs, the directories can't be updated so they begin to classify them subjectively by using centralized learning machine techniques. The writers classify blogs in 4 groups: news, politics, sport and personal. Their precision has been estimated as 84%. In <sup>22</sup>the blogs were classified into two groups: personal and

impersonal. By providing algorithm in compared with data analyses methods such as support vector machine and nearest neighbor of function, they improve the estimation precision and classification. The other group <sup>23</sup>also started to identify political feelings of blog posts and its importance in classification. They used observed effective learning to classify. It can be noted to machine learning algorithms such as Naive Bayes and SVMs. Gerrish et.al<sup>24</sup> began to use co-training to classify blogs by applying political learning. By dividing blogs to two liberal and conservative groups, they stated about the differences of co-training and observed normal education. In this article, Naive Bayes and KNN and its differences with the provided method have been discussed. The other group<sup>25</sup> classified blogs based on the writers' gender. They attach its importance as an issue in business field and used Naive Bayes, regression and support vector machine as new techniques for its improvement. Bhagat et.al<sup>26</sup> pointed out that they encounter object classification in social network analyses. They focus their researches in analyzing all objects classification in a labeled set by using data based on link among the objects. By emphasizing on two algorithm models local and universal similarities, they focus on a formed multi-graph of blog data and apply universal similarities algorithm on some features of bloggers.

### 3. K-Nearest Neighbors algorithm to Classify Bloggers

KNN method was firstly explained in 1950<sup>27,28</sup>. Due to its simplicity, efficiency and application in a few of educational patterns or the same samples, it has been used in recognizing pattern<sup>29,30</sup>. In KNN algorithm, the educational samples are explained by numerical n-dimension features. Each sample is indicated by a point in numerical n-dimension features. So, all the educational samples are saved in an educational pattern of n-dimension. When there is an unknown sample which need classification, this algorithm will look for an experimental pattern for K sample which is nearest to the un-known sample. This approach (nearness) is defined by Euclidean distance. If the two point include  $X_1 = (x_{11}, x_{12}, x_{13}, \dots, x_{1n})$  and  $X_2 = (x_{21}, x_{22}, x_{23}, \dots, x_{2n})$ , so Euclidean distance of the equation will be as follow<sup>31</sup>:

$$dist(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2}$$

**Equation 1.** Euclidean distance between points<sup>29,31</sup>

As Euclidean distance of points is calculated, the label of k neighbor which is in maximum, will give to the unknown sample and classifying is performed by regulating elements based on Euclidean distance<sup>31,32,33</sup>. By considering 3 neighbors, KNN algorithm (K-NN) have been simulated for 100 data base of Kohkiloye and Boyer Ahmad Province bloggers' dataset<sup>1</sup> by using Weka tools. It was simulated based on pseudo-code 1.

The classification error and precision results are shown in Table (1) in compared with classification method of decision Tree<sup>1</sup>. The precision is the number of cases of the class which classified precisely to the all classified cases belongs to that case. Recall of a case is the number of cases in a sample class which classified precisely to the numbers of that class. ROC curve shows the real positive and false positive rate. The false positive rates are the negative tuples rate which wrongly known as positive and provide for a model. F-measure can be averagely considered as

```

procedure nearest neighbor
1. Begin
2. Initialization Parameters
  blogger = {(DegreeI,
  capriceI,topicI,lmtI,lpssI,pnI), . . . ,
  (Degreeen,capricen,topicn,lmtn,lpssn,pbn) }
  new-blogger = (Degreeen+1,
  capricen+1,topicn+1,lmtn+1,lpssn+1) new
  instance to be classified
3. function(classification)
for each blogger do
  distance (i): compute Between new-blogger with
  blogger(i)
ordered blogger set: order blogger set Depending on
  the distance(i)
k: Select the number of nearest neighbors
top k set: select top k nearest neighbors from ordered
  blogger set
  assigns class (pb) of new-blogger using Top majority
  of class (pb) in top k set
4. End
    
```

**Pseudo-code 1.** pseudo-code of K-NN algorithm plan for bloggers classification.

**Table 1.** The results of K-NN and decision tree to classify bloggers

Method	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
Decision Tree[1]	0.912	0.375	0.838	0.912	0.873	0.811	Yes
K-NN	0.926	0.344	0.851	0.926	0.887	0.849	Yes
Decision Tree[1]	0.625	0.088	0.769	0.625	0.69	0.811	No
K-NN	0.656	0.074	0.808	0.656	0.724	0.849	No

a scale of comprehensiveness and precision<sup>31,34</sup>. The full data are shown in Table 1.

In additional discussion, k & b scores delete the possibility of previous possibilities in response and find an appropriate equation for statistical response. Kappa statistic is also a statistical measurement of inter-rater agreement or inter-annotator agreement which is provided for quality cases. The next discussion is that how the accuracy of prediction methods and classification can be measured. It is used error function concept to find the dimension of performed prediction and real values as well as two index of Mean absolute error and Mean squared error to measure prediction error. It is also applied two indexes of Relative absolute error and relative squared error to calculate relative prediction error of each sample values in compared with prediction error toward average value<sup>31,34</sup>. The additional results of simulation and comparison with classification method of decision tree<sup>1</sup> are shown in Table 2.

In Figure1, ROC curves are indicated to classify bloggers in two groups of professional (YES) and non-professional (NO) by using K-NN algorithm and decision tree<sup>1</sup>. The colored lines indicate curve of each group. In ROC curve, positive percentage of accuracy is plotted against positive percent of inaccuracy.

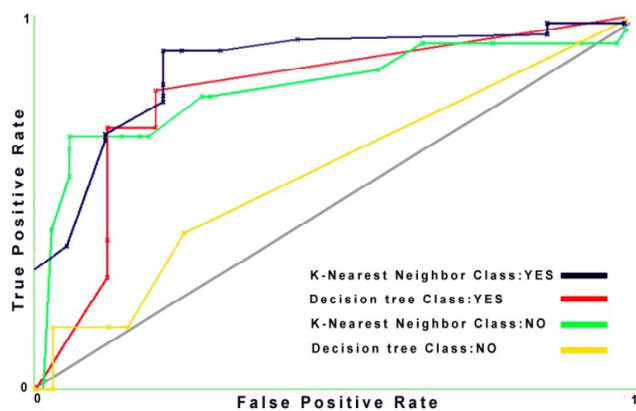
As it can be seen, K-NN has shown better results than decision tree in ROCcurve.

## 4. Artificial Neural Network to Classify Bloggers

In recent decades, methods based on artificial intelligence such as ANNs are used as complement or alternative methods over conventional methods such as algorithmic ones. The neural networks are considered as the member of non-parametric techniques which usually used to estimate and classify<sup>35</sup>. It is demonstrated that the ANNs is more capable in recognizing patterns than the other methods and also is a most used and important tool for classification. The ANNs are more appropriate alternative than the earlier classification methods<sup>36</sup>. The ANNs were brought successful results to classify different test issues such as industry, business, sciences, pattern recognition, sound adaptation, handwriting recognition, medical fields and so on<sup>34,37</sup>. The ANNs are the methods of experimental learning and have demonstrated that are equal or better than the other experimental methods in different fields<sup>38</sup>. The ANNs are firstly developed based on the operation

**Table 2.** The additional results of K-NN algorithm and decision tree precision to classify bloggers

Accuracy	Method	Num	Percent
Correctly Classified instances	Tree[1]	82	82 %
	K-nn	84	84 %
Incorrectly Classified Instances	Tree[1]	18	18 %
	k-nn	16	16 %
Kappastatistic	Tree[1]	0.5648	
	K-nn	0.6132	
K&B Relative Info Score	Tree[1]	5731.732 %	
	K-nn	5898.8665 %	
K&B Information Score	Tree[1]	51.5343 bits	0.5153 bits/instance
	K-nn	53.037 bits	0.5304 bits/instance
Classcomplexity   order 0	Decision Tree[1]	90.5646 bits	0.9056 bits/instance
	K-nn	90.5646 bits	0.9056 bits/instance
Classcomplexity   Scheme	Decision Tree[1]	11831.457 bits	118.3146 bits/instance
	K-nn	81.4824 bits	0.8148 bits/instance
Complexity improvement (SF)	Decision Tree[1]	-11740.8924 bits	-117.4089 bits/instance
	K-nn	9.0822 bits	0.0908 bits/instance
Mean absolute error	Decision Tree[1]	0.1834	-
	K-nn	0.1806	-
Root mean squared error	Decision Tree[1]	0.3876	-
	K-nn	0.3599	-
Relative absolute error	Decision Tree[1]	41.9651 %	-
	K-nn	41.322 %	-
Root relative squared error	Decision Tree[1]	83.0247 %	-
	K-nn	77.076 %	-
UnClassified Instances	Decision Tree[1]	-	-
	K-nn	-	-
Total Number of Instances	Decision Tree[1]	100	-
	K-nn	100	-

**Figure 1.** Roc curve of K-NN algorithm and decision tree precision to classify bloggers.

of human nerves system which includes units called neuron. Their relation in this system determines network behavior. Selecting the type of network is depending on an issue which must be solved. The most common used networks are Back propagation Gradient Network. The major advantages of this network and other ANNs are that they are capable of using potential data. The process of getting data from potential information is called network learning and assigned to determine weights in ANNs<sup>31,39</sup>. Multilayer Feed-Forward is a sample of ANNs in which the learning process of neural networks is done by back propagation algorithm. It is consistently repeated the weight learning process to classify and predict a class of samples. Multilayer Feed-Forward network is formed

of input layer, one or more hidden layer and output layer. Each layer is formed of units. The network entry is the scale of features which have learning sample and are entered entry or input layer simultaneously. Then, they became weighted and transferred to the second layer of pseudo-neuron units which known as hidden layers. The output of hidden layer can entered as an entry to the other hidden layer. The weighted output layer or hidden layers are used as the entry of output layer which can provide classification process and prediction of samples and test samples by reaching to ideal weight after network learning of educational dataset<sup>40</sup>. In pattern recognition issues, we need an artificial neural network to classify entry or input to a set of goal groups. The ANNs pattern of recognition tool will help us to select data, create and provide network learning and perform our performance assessment by using mean square error and confusion matrix. In this paper, we use not only K-NN method to classify bloggers but also ANNs to increase accurate classification rate. As a result, by creating and the ANNs learning, we begin to assess performance by applying precise classification indexes.

Firstly, we create an artificial neural network by using MATLAB tools. Our artificial neural network is a feed forward network with tan-sigmoid transmission function in output and hidden layers it is used 10 neurons in the hidden layer. The network includes 5 inputs and 2 outputs as the goal vector has 2 members and allocates to 2 groups of bloggers (professional and non-professional) for 100 dataset of Kohkiloye and Boyer Ahmad Province<sup>1</sup>. The general architecture of feed-Forward artificial neural network to classify bloggers is shown in Figure 2.

The number of input nodes in ANNs, the number of nodes in hidden layer and the number of output nodes are usually arranged by final data, error and test and final decision parameters, respectively<sup>41</sup>. Learning network, test and artificial neural network display are shown in pseudo-code 2.

We set the number of neuron in the hidden layer equal to 10 and create the network on the basis of 2 hidden layers. In the other stage, all the input data are classified in groups randomly. We use 70, 10 and 20 % of data for learning program, validation and network test, respectively. In network learning section, by selecting Gradient learning and scaled function (as learning function), we began to network learning.

It is necessary to note that by using NPRTOOL in MATLAB tools, the efficiency of network will review as

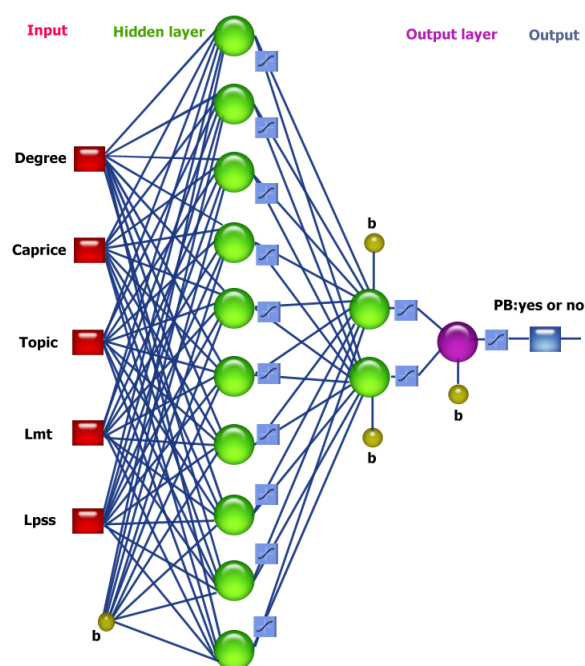


Figure 2. The general architecture of feed-Forward artificial neural network to classify bloggers.

**Procedure neural networks**

1. **Begin**
2. **Initialization Parameters**  
 inputs:  $\{(Degree1, caprice1, topic1, lmt1, lpss1), \dots, (Degree_n, caprice_n, topic_n, lmt_n, lpss_n)\}$   
 targets:  $\{(pb:yes1 \dots pbyes_n), (pb:no1 \dots pbno_n)\}$
3. **Hidden Layer Size**  
 Select the number of hidden layers;
4. **Net**  
 pattern net(hidden Layer Size);
5. **net. Divide Fcn**  
 Choosing how to divide the sample
6. **Net. Divide Mode**  
 Choosing mode of divide the sample;
7. **SELECT type of sample**
8. **Net. Train Fcn**: conjugate gradient back propagation  
 selected Network training function
9. **Net. Perform Fcn**: Mean squared error  
 Choose a Performance Function
10. **Train the Network**  
 $[net, tr] = \text{train}(net, inputs, targets);$
11. **Test the Network**  
 output = net(inputs);  
 compute error of network
12. **Performance**  
 compute performance of network
13. **End**

Pseudo-code 2. pseudo-code of planning artificial neural network to classify bloggers.

default by Mean squared error and the learning process will be stopped, if the assessment error increased during 6 consecutive and repeated stages. For this issue, it is happened in 45 repetitions. The diagram of learning errors, assessment errors and test errors are shown in Figure 3.

As it can be seen, the best efficiency is achieved in 39 repetitions. The learning procedure of learning data is achieved 45 repetitions and shown in Figure 4.

In Figure 5, the histogram of learning stages errors are shown.

To analyze network response of confusion matrix and different errors in final network case is shown in Figure 6.

Diagonal elements in each table indicate the number of cases which classified accurately. The non-diagonal elements are those which classify inaccurately. The blue-colored cell (in the lower right corner of depicted tables in

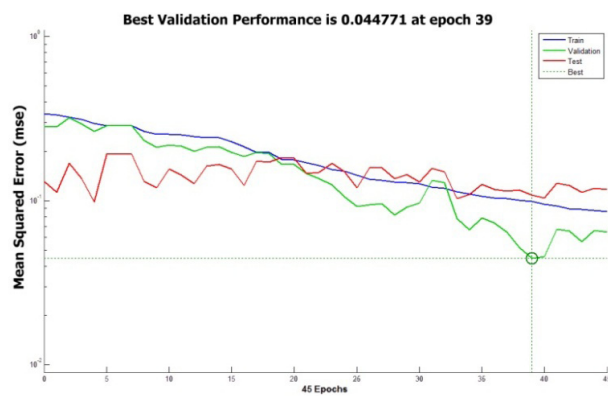


Figure 3. The diagram of learning errors, (validation) errors and test errors of feed-Forward artificial neural network to classify bloggers.

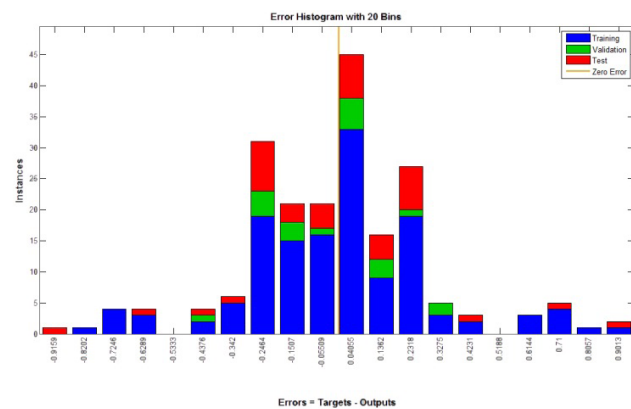


Figure 4. The learning procedure of learning data of feed-Forward artificial neural network to classify bloggers.

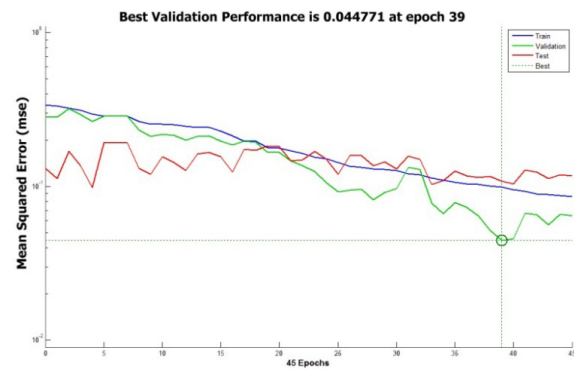


Figure 5. The histogram of learning stages errors of feed-Forward artificial neural network to classify bloggers.



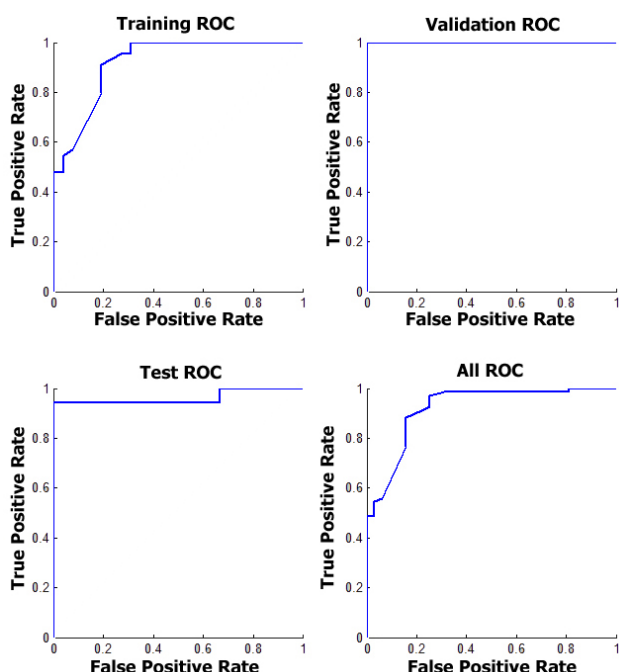
Figure 6. Confusion matrix and Feed-Forward artificial neural network errors to classify bloggers.

Figure 6) indicates total percentage of inaccurate classification in red color. In Figure 7, ROC curve are indicated for learning stage, validation, test and finally 3 stages inclusion in general. In ROC curve, positive percentage of accuracy is plotted against positive percent of inaccuracy.

As it can be seen in ROC curve, the artificial neural network has appropriate function in classification.

## 5. Discussion and Evaluation

In Lazy method of classification all learning samples are saved firstly and no classification is performed till the unknown sample didn't need it. It is completely opposite



**Figure 7.** ROC diagram of Feed-Forward artificial neural network to classify bloggers.

of eager methods. In eager methods, a model is created by using learning data before unknown sample entry. As it enters, it is classified by using previously created model<sup>29,32</sup>. The difference of these two methods is that in eager method, a lot of time is spent to create model in learning stage as the classification is performed very quick. But in lazy method, the data and learning network has saved the model at first and then, spent more time to classify<sup>34,42,43</sup>. Decision Tree<sup>1</sup> and artificial neural network methods were firstly created a model by using learning data and then used the same previous model for new entries or inputs. So, they are considered as eager method. However, K-NN algorithm is a method based on the nearest learning samples in features space. And, as the new sample entered to classify, the model will be created. So, it is the sample of lazy method. The types of classification of these methods to classify bloggers are shown in Table 3.

There are also differences in the precision of classification in 3 different methods as it can be seen in the table. In Table 4, the rate of accurate classification of provided methods are indicated based on decision Tree method<sup>1</sup>.

Considering the results of Feed-Forward algorithms, K-NN and decision Tree<sup>1</sup> methods to classify bloggers, it can be seen that K-NN method improved 2% the classification results of decision Tree<sup>1</sup>. Then, by providing artificial

**Table 3.** The comparison of Feed-Forward ANNs algorithms, KNN and decision Tree classification methods

Method	Classification type
Decision Tree[1]	Eager classification
K-NN	Lazy classification
ANN	Eager classification

**Table 4.** The comparison of the precision of Feed-Forward artificial neural network algorithms, K-NN and decision Tree to classify bloggers

Method	Accuracy %
Decision Tree[1]	82
K-NN	84
ANN	90

neural network, K-NN classification results improved, too. And finally, to get the precision to 90% accuracy, by using artificial neural network, we increase precision to 6 and 8 % toward K-NN algorithm and Decision Tree, respectively.

## 6. Conclusion

Web logs are one of the effective tools of web2 which has more capabilities than other web2 tools. The bloggers are those administrated blogs and can be classified either professional or non-professional. The tendency of the bloggers can be determinant of their social characteristics and living environments. So, bloggers classification is considered important. Two methods were used in this paper: KNN and Artificial Neural Network to classify bloggers of Kohkiloye and Boyer Ahmad province in Iran. The simulation test results have demonstrated that the provided method had relative improvements over earlier approaches and provided better results for bloggers to classify.

## 7. References

- Gharehchopogh FS, Khaze SR. Data mining application for cyber space tendency in blog writing: a case study. IJCA. 2012; 47(18):40–6.
- Chang C-F. Exploring genres and mediated actions in Taiwanese college students' blog writing. The JALT CALL Journal. 2011; 7(2):137–50.
- Tekinarslan E. Blogs: a qualitative investigation into an instructor and undergraduate students' experiences. Australas J Educ Tech. 2008; 24(4):402–12.



4. Abidin MJZ, Pour-mohammadi M, Hamid FBA. Blogging: promoting peer collaboration in writing. *Int J Bus Humanit Tech.* 2011; 1(3):98–105.
5. Karandikar A. Generative model to construct blog and post networks in blogosphere [M.Sc Thesis]. University of Maryland; 2007.
6. Sun Y-C, Chang Y-J. Blogging to learn: becoming EFL academic writers through collaborative dialogues. *Lang Learn Tech.* 2012; 16(1):43–61.
7. Arnot K. Blogging the Gap: A survey of China bloggers [M.Sc Thesis]. London School of Economics and Political Science; 2010 Aug.
8. Lin F, Cohen WW. The multi rank bootstrap algorithm: self-supervised political blog classification and ranking using semi-supervised link classification. *ICWSM.* Seattle, Washington, USA; 2008.
9. Ines B, Nicolas B, Mathieu R. Blog classification: adding linguistic knowledge to improve the K-NN Algorithm. *Intelligent Information Processing 4.* Springer; 2008. p. 68–77.
10. Jun Y, Park H, Myaeng SH. A hybrid mood classification approach for blog text. *Trends in Artificial Intelligence.* 2006; 4099:1099–103.
11. Sriphaew K, Takamura H, Okumura M. Cool blog classification from positive and unlabeled examples. *Advances in Knowledge Discovery and Data Mining.* Springer; 2009. p. 62–73.
12. Wiegand M, Klakow D. Topic-Related polarity classification of blog sentences. *Progress in Artificial Intelligence.* Springer; 2009. p. 658–69.
13. Rustagi M, Prasath RR, Goswami S, Sarkar S. Learning age and gender of blogger from stylistic variation. *Pattern Recognition and Machine Intelligence.* Springer; 2009. p. 205–12.
14. Yang S, Yan J, Gao C, Tan G. Blogger's interest mining based on chinese text classification. *Nonlinear Mathematics for Uncertainty and its Applications.* Springer; 2011. p. 611–8.
15. Ishino A, Nanba H, Takezawa T. Automatic classification of link polarity in blog entries. *Information Retrieval Technology.* Springer; 2011. p. 509–18.
16. Ayyasamy RK, Alhashmi SM, Eu-Gene S, Tahayna B. Enhancing concept based modeling approach for blog classification. *Knowledge Engineering and Management, Part 5.* Springer; 2012. p. 409–16.
17. Ayyasamy RK, Alhashmi SM, Eu-Gene S, Tahayna B. enhancing automatic blog classification using concept-category vectorization. *Knowledge Engineering and Management, Part 5.* Springer; 2012. p. 487–97.
18. Tsai FS. Blogger-Link-Topic model for blog mining. *New Frontiers in Applied Data Mining.* Springer; 2012. p. 28–39.
19. Yan X, Yan L. Gender Classification of Weblog. *Proceedings of AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs.* 2006. p. 228–30.
20. Sun A, Suryanto MA, Liu Y. Blog classification using tags: an empirical study. *Asian Digital Libraries.* Springer; 2007. p. 307–16.
21. Qu H, Pietra AL, Poon S. Automated Blog Classification: Challenges and Pitfalls. *Proceeding of AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs.* 2006. p. 184–6.
22. Elgersma E, de Rijke M. Personal vs non-personal blogs: initial classification experiments. *SIGIR.* New York, USA: ACM; 2008. p. 723–4.
23. Durant KT, Smith MD. Predicting the political sentiment of weblog posts using supervised machine learning techniques coupled with feature selection. *Philadelphia, USA: WEBKDD; 2006.* p. 187–206.
24. Gerrish J, Qazvinian V, Shi X. Blog Classification with Co-training. 2007.
25. Mukherjee A, Liu B. Improving Gender Classification of Blog. *EMNLP; 2010.* p. 207–17.
26. Bhagat S, Rozenbaum I, Cormode G. Applying link-based classification to label blogs. *WebKDD/SNA-KDD.* New York, USA: ACM; 2007. p. 92–101.
27. Fix E, Hodges JL. Discriminatory analysis, nonparametric discrimination: Consistency properties. *Technical Report 4.* Randolph Field, Texas: USAF School of Aviation Medicine; 1951.
28. Silverman BW, Jones MC, Fix E, Hodges JL. An important contribution to nonparametric discriminant analysis and density estimation: commentary on fix and hodges. *International Statistical Review.* 1989; 57(3):233–8.
29. Phyu T. Survey of Classification Techniques in Data Mining. *International Multi Conference of Engineers and Computer Scientists.* 2009 Mar; Hong Kong.
30. He J, Tan A, Tan C. Comparative study on chinese text categorization methods. *The PRICAI 2000 Workshop on Text and Web Mining.* 2000. p. 25–31; Melbourne.
31. Han J. *Data Mining: Concepts and Techniques.* Morgan Kaufmann; 2006. p. 348–50.
32. Klair AS, Kaur RP. Software effort estimation using SVM and KNN. *International Conference on Computer Graphics, Simulation and Modeling.* 2012 Jul. p. 146–7; Thailand.
33. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou Z-H, Steinbach M, Hand DJ, Steinberg D. Top 10 algorithms in data mining. *Journal Knowledge and Information Systems.* 2007 Dec; 14(1):1–37.
34. Gorunescu F. *Data mining concepts, models and techniques.* Intelligent Systems Reference Library. Springer; 2011. p. 256–60.

35. Gharehchopogh FS. Neural networks application in software cost estimation : a case study. International Symposium on Innovations in Intelligent Systems and Applications (INISTA); 2011; Istanbul. p. 69–73.
36. Gharehchopogh FS, Ahmadzadeh E. Artificial neural network application in letters recognition for farsi/arabic manuscripts. International Journal of Scientific and Technology Research. 2012 Sept; 1(8):90–4.
37. Zhang GP. Neural networks for classification: a survey. IEEE Trans Syst Man Cybern C Appl Rev. 2000 Nov; 30(4):451–62.
38. Reby D. Artificial neural networks as a classification method in the behavioral sciences. Behavioral Processes. 1997; 40: 35–43.
39. Svozil D. Introduction to multi-layer feed-forward neural networks. Chemometr Intell Lab Syst.1997; 39:43–62.
40. Towell GG, Shavlik JW. Knowledge-based artificial neural networks. Artif Intell.1994; 70(1-2):119–65.
41. Gharehchopogh FS, Khalifelu ZA. Neural network application in diagnosis of patient: a case study. International Conference on Computer Networks and Information Technology (ICCNIT); 2011; Abbottabad. p. 245–9.
42. Martin B. Instance-based learning: nearest neighbor with generalization. University of Waikato; 1995.
43. Larose DT. Discovering knowledge in data: an introduction to data mining. Wiley-Interscience; 2005.