

# News recommendation system using machine learning

Neha Rani <sup>1\*</sup>, Sudhir Pathak <sup>1</sup>

<sup>1</sup> Chandigarh University, India

\*Corresponding author E-mail: [nhjasrotia.24@gmail.com](mailto:nhjasrotia.24@gmail.com)

## Abstract

The forecasting of financial news is yet becoming the main issue to divide the new into different classes on the basis of present time series. Moreover, it might be utilized for predicting and analyzing the stock market for the particular industry. Thus, the new content is significantly important to influence market forecast report. In this paper, the financial news from four countries namely America, Australia, India and South Africa along with their stop words are consider. The words along with their weighted values are determined and then the neural network is trained. Here, artificial neural network is used for classifying the appropriate results for the given input data. At last the comparison of ANN with SVM is shown. Experiments show that the ANN classification provides high accuracy to predict the news than the SVM classifier.

**Keywords:** News Classification; ANN; SVM; Recommendation System.

## 1. Introduction

Nowadays, the way of reading news has changed due to the use of the Internet. News providing websites such as Google & Yahoo gathers the news from different sources and presents an aggregate representation of news throughout the world. A significant difficulty with news providing websites is that the quantities of articles can be overpowering to the users. Thus, it becomes a challenge to search news article, which is of user's interest [1].

The solution to this problem is to design a content based recommendation system. The content based recommendation system is

used to resolve the problem of overload. The system categorized the items as per the user's interest. This system can be used in number of domains such as email, news and web search. In the field of news, this system primarily aims at gathering news articles as per user interests and forming a personal newspaper for a particular user. The recommendation system is mainly categorized into two types named as Content based and collaborative filtering algorithm [2]. Figure 1 shows collaborative filtering and content based filtering.

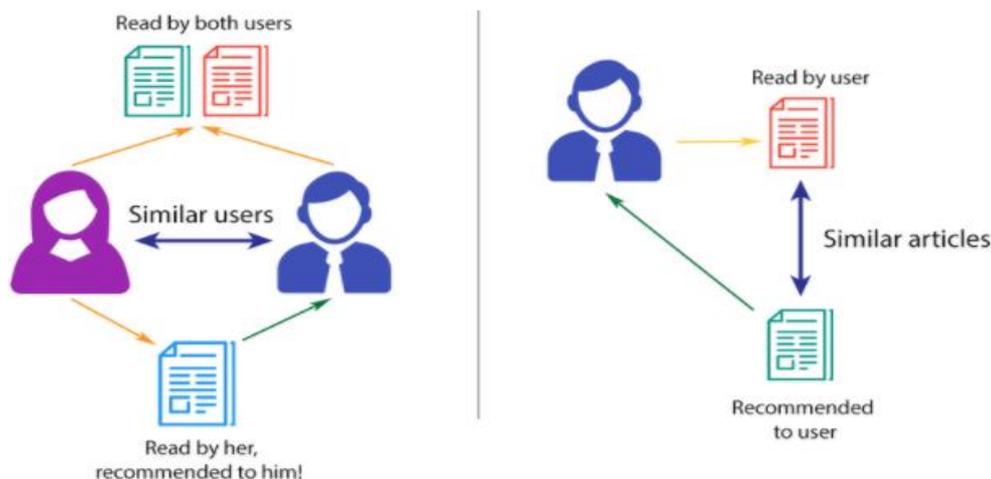


Fig. 1: Collaborative Filtering and Content Based Filtering.

### 1.1. Content based algorithm

This type of recommendation system works with the user provided data, which is provided by means of rating and clicking on the website link. A user profile is created on the basis of available

data and used for making suggestions to the client [3]. With the increase in the number of inputs, the recommendation system becomes more accurate. The relative importance of the news articles is determined by using TF (Term frequency) and IDF (Inverse document frequency). The TF and ITF of the news article can be

measured by using the log function which is given by the equation 1.

$$W_{x,y} = \left\{ \begin{array}{ll} 1 + \log_{10} tf_{xy}, & \text{if } tf_{xy} > 0 \\ 0 & \text{Otherwise} \end{array} \right\} \quad (1)$$

The similarity of the news article is measured by the equation (2) written below.

$$S(Q_j, Q_k) = \frac{Q_{\sim j, Q_{\sim k}}}{|Q_{\sim j, Q_{\sim k}}|} \quad (2)$$

Here  $W_{x,y}$  represents the weight of news articles [4]

## 1.2. Collaborative filtering

The articles are provided to the user on the share basis. If the two different clients have the same choice, a group is created. Hence, it is also named as user based system/ an item based system [5].

## 2. Related work

Liu et al. [6] proposed 'news recommendation system' used in Google news. The system makes dataset according to the number of clicks provided by the user's interest. A framework named as Bayesian has been designed as per the activities shown by the users. Content based along with collaborative filtering algorithms are used in hybridization. Quality of the 'news recommendation system' has been enhanced. It covers entire items of the user's interest. The system upgraded automatically. The parameters such as click through rate, frequency of Google news website visit/ day have been measured.

Li et al. [7] proposed a two phase news recommendation system that comprises of news data during the performance of recommendation. The authors used the intrinsic properties of the user choice that have a better balance among the originality and multiplicity of the recommended outputs. The data has been gathered from various news articles. K-mean clustering algorithm has also been used. The parameter named as F-score has been measured.

Chidambarathanu et al. [8] proposed SVM (Support vector machine) classification approach that assists the clients to create informed communication. Classifier used the social media data so that accurate decision has been takes place. Comparison of traditional prediction system along with SVM based system is provided and concluded that SVM has provided better accuracy than traditional approach.

Li et al. [9] presented a news article recommendation system to solve the problem of contextual bandit. In this method a learning system automatically chose the articles on the basis of user's contextual information and the user's past clicking options. Three proposes has been performed, first is the proposed work is computationally effective, second is the bandit approach has been evaluated without internet by using past recorded random and lastly the proposed approach is applied to Yahoo.

Demirsoz, O., and Ozcan, R. [10] proposed to form a training set using tweets containing a link to the news and the content of the same news article. The tweets have been preprocessed by removing the undesired words and symbols. The morphological process has been applied for stemming along with anomaly detection.

## 3. Materials and methods

In this research work, optimization algorithm, Artificial Bee colony scheme is used to provide optimal feature set from the text document depends upon the objective function. Also a classification algorithm, ANN (Artificial neural network) is used to classify the multiple categories of news as per the user's choice. The algorithms of ABC and neural networks are defined below.

### 3.1. Artificial bee colony (ABC)

ABC is an optimization algorithm used to find the most appropriate solution by using the intellectual behavior of honeybees. In this research work, ABC is used for the selection of news features. Employed bees, onlooker bees' and scout bees' are the three main groups in the ABC algorithm [11]. Figure 2 shows proposed work flowchart.

#### Algorithm 1: Artificial Bee colony

```

Step 1: Load the samples for training
Step 2: Create initial population P, t=1.....N
Step 3: Initialize the bee categories- Employed bee
- Onlookers bee and
- Scouts bee
Step 4: Compute the population's fitness function (ft)
Step 5: Put j= 1
Step 6: Repeat
Step 7: For every employee Bee{
Create novel solution yt by using step 7
Compute the fitness value
Use greedy selection approach}
Step 8: Determine the probability values P for obtaining the result by
step 6
Step 9: For every onlooker bee {
Choose a solution on the basis of probability P
Generate new solution yt
Determine the fitness value
Apply 'Greedy selection approach'}
Step 10: If the solution for the scout is abandoned
Then replace it with a novel solution that will be generated randomly by
step 8
Step 11: Learn the better solution so far
Step 12: j=j+1

```

### 3.2 Artificial neural network

Artificial neural network is used to classify the news categories by using the biological behaviour of the neurons. ANN consist number of neurons that are interconnected with each other. Bias is used to adjust the value of neurons so that desired output can be obtained [12].

#### Algorithm 2: Neural Network

```

Input: Train n number of documents
Output: Classified document
Divide the training group into training 'S' and validation 'T'
Calculate the accuracy of classifier
Define parameters
Training document=optimized document
Group= similar news
Epoch=100
Training algo= Levenberg Marquardt
Performance metrics= Mean square error
Neurons=50
Net-neuron (Trained_data, group, neurons)
Classification={News as per choice,; if properties are matched
Return {classified news}

```

### 3.3 Support vector machine

In this research work, SVM is used for the classifying the text among the new articles. It is utilized to know the interest of users on internet and provide guidelines to the users through the hypertext. The input is provided to the SVM and with the help of Kernel, the documents are mapped into the high dimensional space [13]. SVM distinguished between the similar and non similar documents by using hyper plane structure the algorithm is listed below:

**Algorithm 3: Support Vector Machine**

```

Input: Initialize Properties of document
Output: Classified document
Initialize the SVM training data 'X'
Defined 'C' as a class of training data
Set the  $f_k$  as Kernal function
For each news
Train_data=SVMTRAIN (X,C,  $f_k$ );
End
News interest list=SVMCLASSIFY (Train_data,test properties of test document)
Classify news
Else
Check it again and return;
End
End function
    
```

**Training phase**

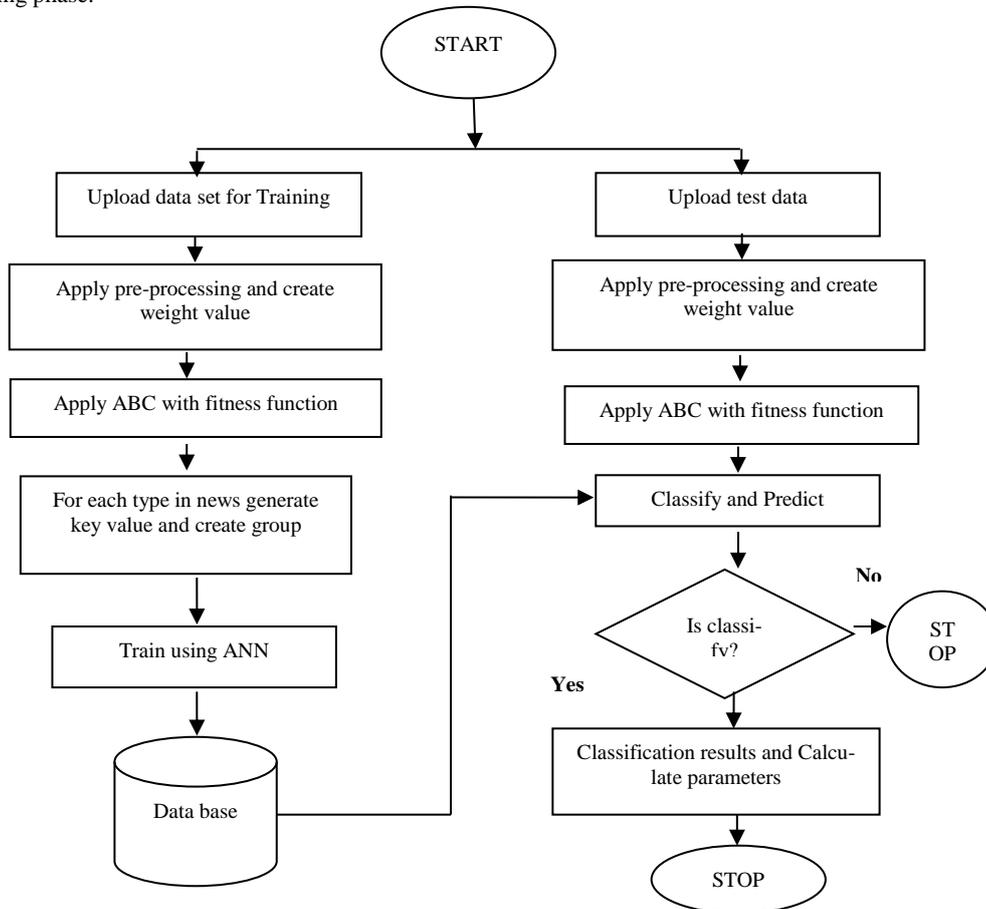
- i) In training phase, pre-processing is applied on the uploaded document and their weighted values are generated.
- ii) ABC algorithm is applied to optimize the features of the loaded document by using the fitness function.
- iii) For every type of news unique key values are generated and the values that have same weight form the group.
- iv) Neural network is trained on the basis of generated key values.

**Testing phase**

- i) Test document is loaded to know the proficiency of the designed recommendation system.
- ii) Apply pre-processing and generate the weighted values
- iii) Optimization is applied on the generated weighted values and then the values are classified by comparing with the data base values stored into the ANN database.
- iv) If the values are matched then uploaded document is classified and QoS of the proposed design are evaluated.

The procedure of simulation of the proposed work is defined below:

The proposed research work is processed in two phases named as training and testing phase.



**Fig. 2:** Proposed Work Flow Chart.

**4. Simulation results**

References [14] – [25] have inspired this research. The experiment is performed in MATLAB simulator. The performance of prediction system is measured in terms of precision, recall, true positive and false positive parameters. ANN and SVM classifier are used for news classification. The comparison of the classification techniques are also provided to know the accuracy of the classifiers.

Figure 3 defines the GUI of the proposed work. The GUI consists of two panels named as Community profiles and test data. The community panel is categorized into eight different sub parts named as Community1, Community2, Community3, and Community4, train the system, neural network and identify the community. The recognized data is shown under Test data block.

Figure 4 depicts the community uploaded along with their corresponding weights. The community along with their word and weight are shown in figure above.

After clicking on the “Train the system button” the above Figure 5 displays on the screen. There are 14 number of neurons provided at the input layer. These neurons are passed to the hidden layer. In hidden layer, 6 numbers of neurons are added as a bias, so that the desired output can be obtained. At the output layer of neural network 1 neuron is obtained.

Figure 6 represents the training state of the neural network. The parameters such as gradient, mutation and validation check are measured. The neural network is trained for four number of iterations and the value of gradient, mutation and validation check are 1.5952, 1e-07 and 4 respectively.

Figure 7 represents the regression state of neural network. The training, validation, test and the regression values are shown in the form of graph.

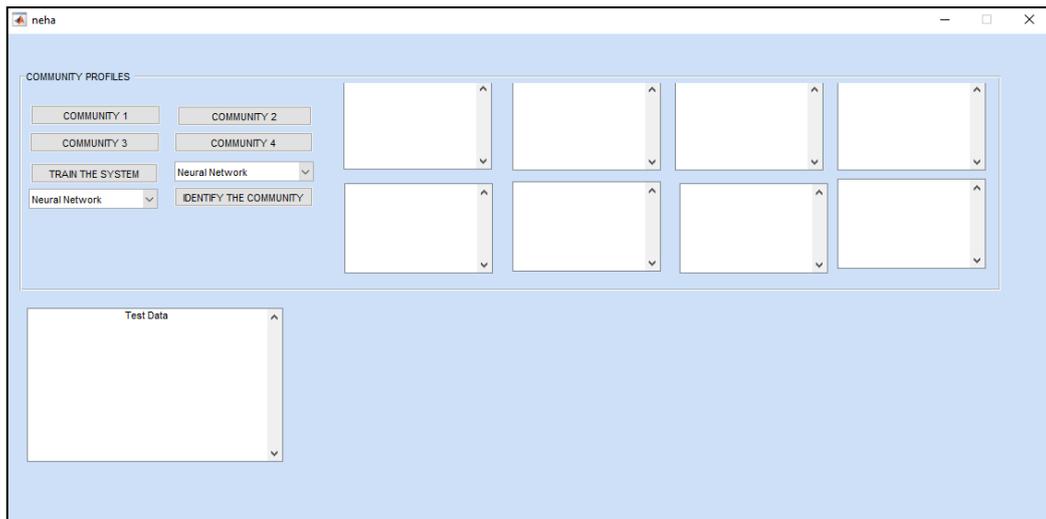


Fig. 3: GUI of Proposed Work.

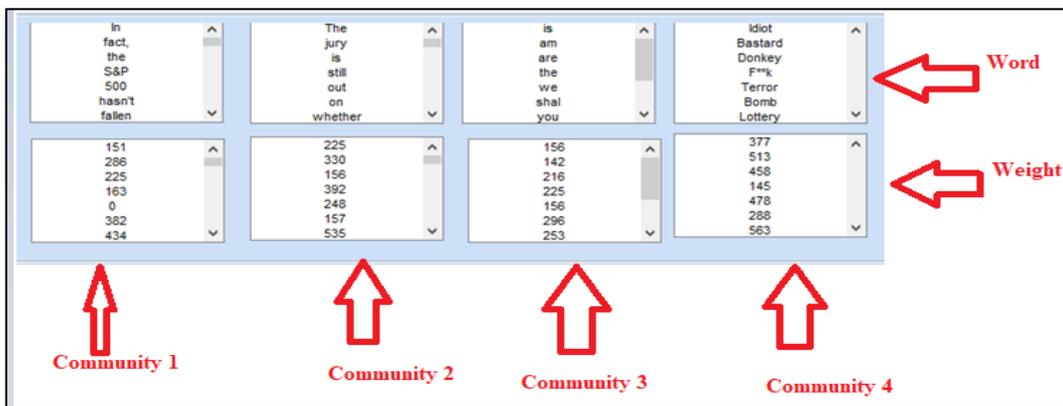


Fig. 4: Word and Weight of Uploaded Test Community.

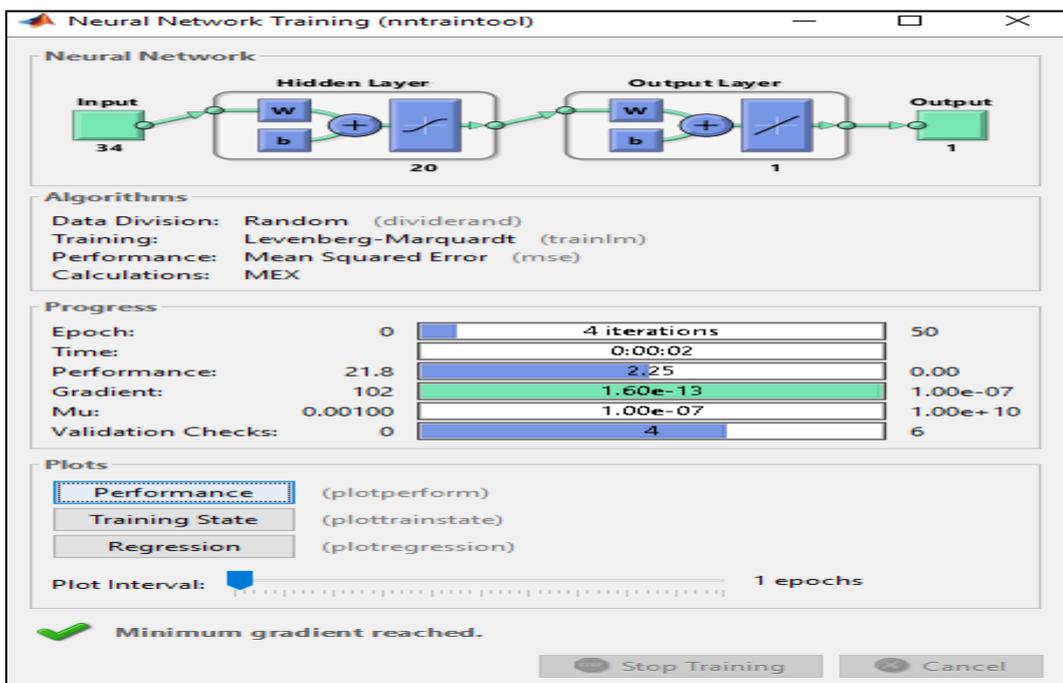


Fig. 5: Performance of ANN.

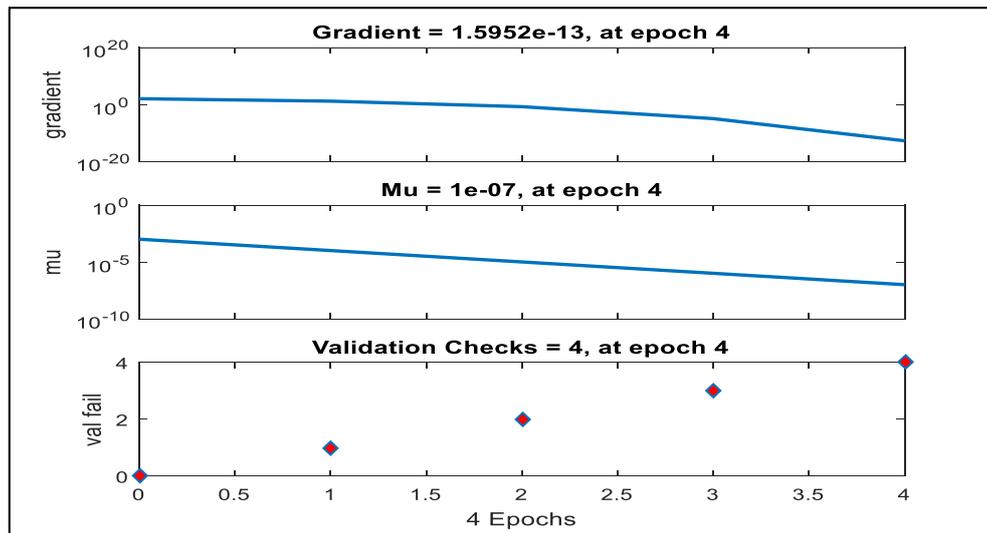


Fig. 6: Training State.

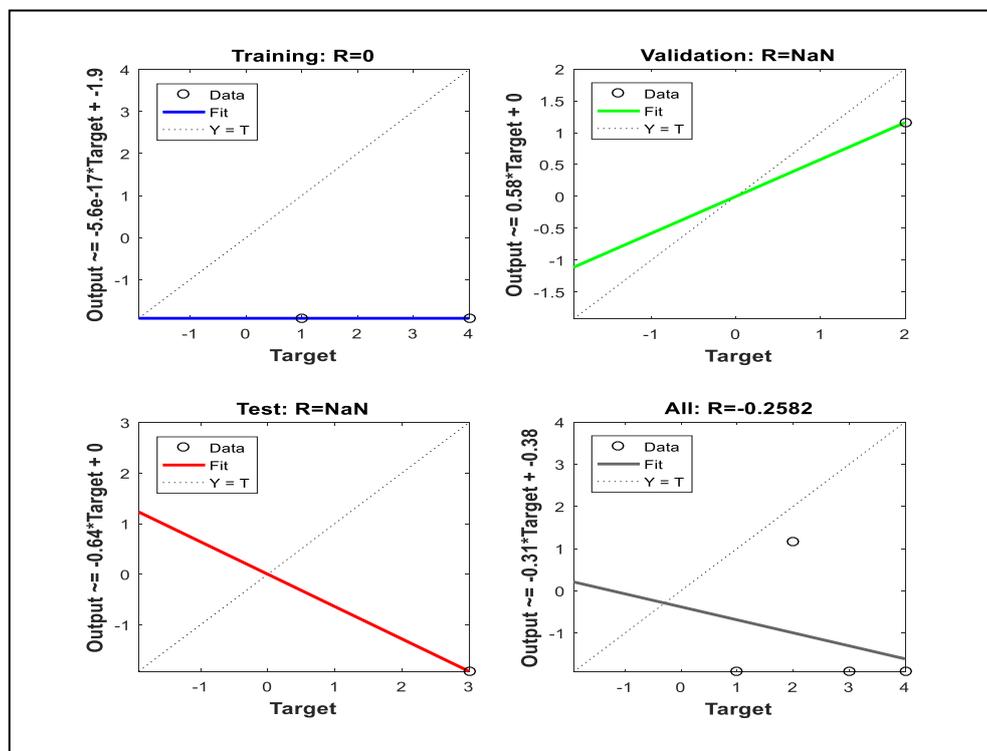


Fig. 7: Regression State.

Figure 8 shows SVM Classification. The SVM plot obtained for community 1, community 2, community 3 and community 4 are shown in figure (a), (b), (c) and (d) respectively. Here, the straight line represents the kernel, red colour with plus sign represents the elements and green colour represents the community and the green colour encircle with black colour represents the classified group. The parameters such as precision, recall, and true positive and false positive are measured to know the effectiveness of the proposed work. The parameters observed for 10 numbers of iterations are listed in Table 1 and Table 2 below.

Figure 9 depicts the comparison of precision and recall obtained for the proposed work by applying artificial neural network (ANN) and Support Vector Machine (SVM). Here, Blue and red bar graph represents the precision and recall values for ANN whereas green and purple bar graph represents the precision and recall values obtained for the SVM classifier. The code is run for 10 numbers of iterations and the average values obtained for precision and recall when ANN is applied are 1.364 and 1.328 respectively. After applying SVM, the average values obtained for precision and recall are 1.322 and 1.304 respectively.

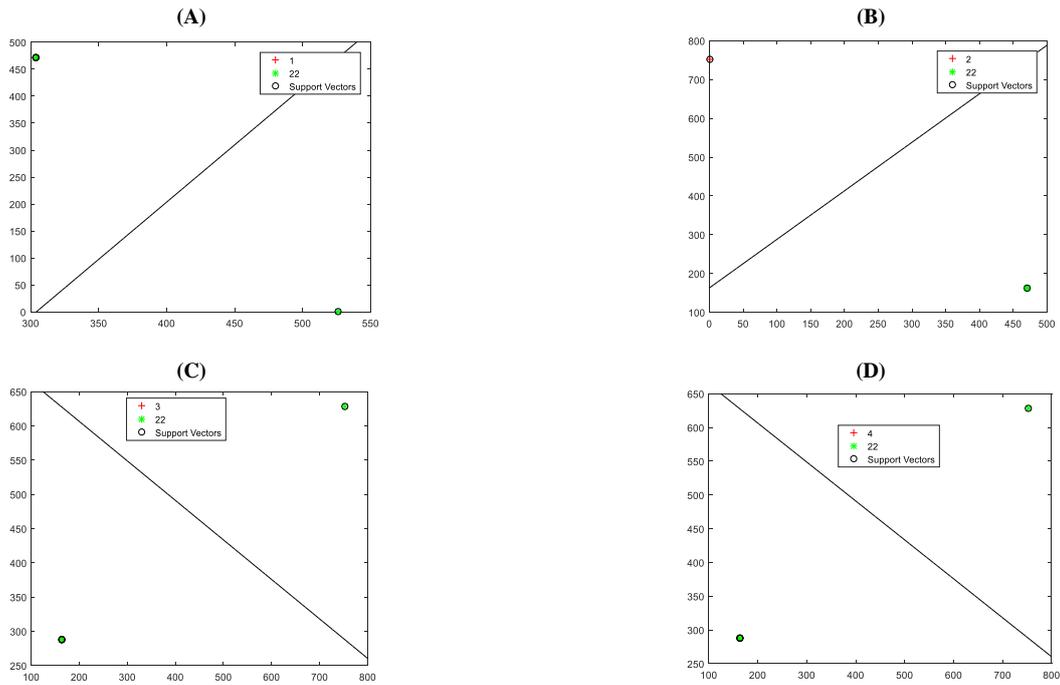


Fig. 8: SVM Classification.

Table 1: Parameters Using Neural Network

Number of iterations	Precision	Recall	True positive	True negative
1	1.4	1.38	0.0007987	0.00058393
2	1.3	1.29	0.0007786	0.00057964
3	1.34	1.35	0.0007689	0.00058354
4	1.29	1.26	0.0007980	0.00058378
5	1.28	1.27	0.0008081	0.00054752
6	1.45	1.31	0.0008075	0.00056872
7	1.32	1.24	0.0007824	0.00055427
8	1.39	1.37	0.0007821	0.00054536
9	1.46	1.41	0.0007829	0.00059687
10	1.41	1.40	0.0007975	0.00059257

Table 2: Parameters Using SVM

Number of iterations	Precision	Recall	True positive	True negative
1	1.32	1.37	0.0006859	0.00068522
2	1.25	1.25	0.0007666	0.00059672
3	1.31	1.31	0.0006587	0.00059687
4	1.25	1.24	0.0006980	0.00060254
5	1.24	1.26	0.0007081	0.00064578
6	1.42	1.27	0.0007175	0.00059687
7	1.30	1.23	0.0006824	0.00056358
8	1.35	1.36	0.0007121	0.00057021
9	1.40	1.40	0.0006229	0.00062487
10	1.38	1.35	0.0006975	0.00061235

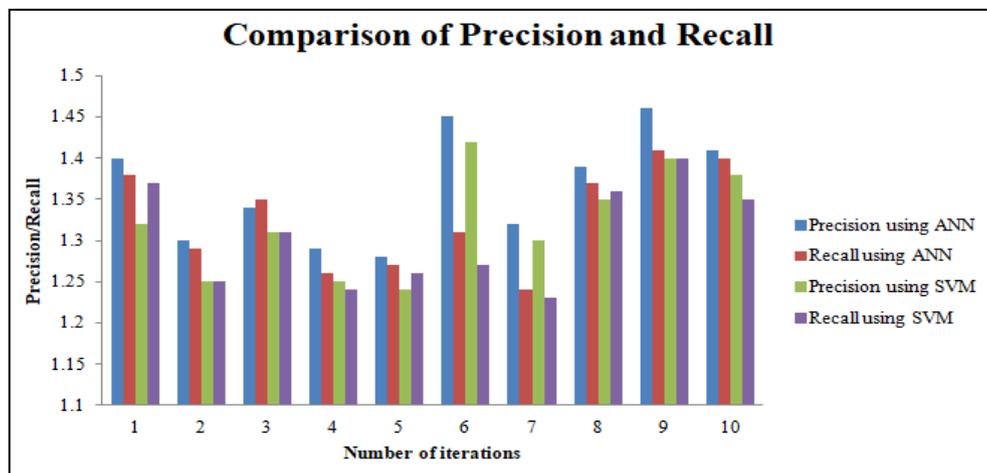


Fig. 9: Comparison of Precision and Recall.

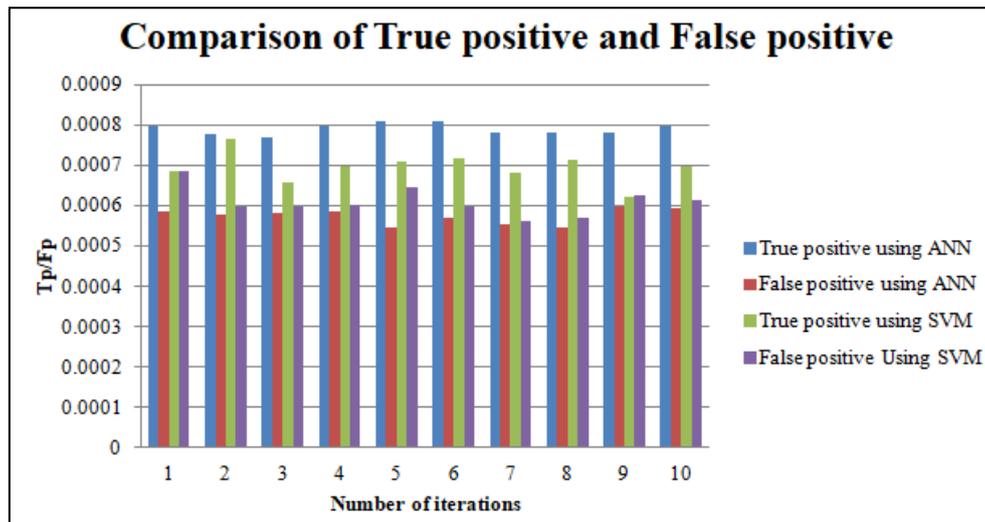


Fig. 10: Comparison of True Positive and False Positive.

The above Figure 10 defines the comparison of true positive and false positive measured for the proposed work by using ANN and SVM classifier. Here, x axis represents the number of iterations and y-axis represents the values of true positive (Tp) and false positive (Fp) values. The average values of Tp and Fp obtained for 10 number iterations when ANN is applied are 0.0079047 and 0.00057362 respectively. The average values of Tp and Fp for SVM classifier are 0.00069497 and 0.000609501 respectively.

## 5. Conclusion

In this research work, financial news are classified on the basis of content of appropriate news articles that has been accomplished by designing a prediction model that is capable to classify the news. The stock market has been studied repeatedly to extract valuable information and estimate their movement. Economic news only based upon technical and basic data classification analysis. Textual data such as news editorial has better information, therefore, utilizing textual information can enhances the input to the data of numeric time series and the better predications are imagined from this type of input than only from statistical data. It has been concluded that the artificial neural network perform better as compared to the SVM classifier. To know the efficiency of the prediction system, the parameters named as precision, recall, true positive and false positive have been measured. The average values obtained for the precision and recall by using ANN is 1.364 and 1.328 respectively. After applying SVM, the average values obtained for precision and recall are 1.322 and 1.304 respectively. The value of Tp and Fp, when ANN is applied are 0.0079047 and 0.00057362 respectively. The average values of Tp and Fp for SVM classifier are 0.00069497 and 0.000609501 respectively. The ANN classified the news with higher accuracy than the SVM. This is because SVM is a binary classifier whereas ANN is a multilevel classifier.

## References

- Trieu, L. Q., Tran, H. Q., & Tran, M. T. (2017, December). News Classification from Social Media Using Twitter-based Doc2Vec Model and Automatic Query Expansion. In *Proceedings of the Eighth International Symposium on Information and Communication Technology* (pp. 460-467). ACM. <https://doi.org/10.1145/3155133.3155206>.
- Buhl, F., Günther, E., & Quandt, T. (2018). Observing the dynamics of the online news ecosystem: News diffusion processes among german news sites. *Journalism Studies*, 19(1), 79-104. <https://doi.org/10.1080/1461670X.2016.1168711>.
- Xia, Y., Di Fabrizio, G., Vaibhav, S., & Datta, A. (2017). A Content-based Recommender System for E-commerce O ers and Coupons.
- Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender systems handbook* (pp. 73-105). Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-85820-3\\_3](https://doi.org/10.1007/978-0-387-85820-3_3).
- Wei, J., He, J., Chen, K., Zhou, Y., & Tang, Z. (2017). Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications*, 69, 29-39. <https://doi.org/10.1016/j.eswa.2016.09.040>.
- Liu, J., Dolan, P., & Pedersen, E. R. (2010, February). Personalized news recommendation based on click behavior. In *Proceedings of the 15th international conference on Intelligent user interfaces* (pp. 31-40). ACM. <https://doi.org/10.1145/1719970.1719976>.
- Li, L., Wang, D., Li, T., Knox, D., & Padmanabhan, B. (2011, July). SCENE: a scalable two-stage personalized news recommendation system. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 125-134). ACM. <https://doi.org/10.1145/2009916.2009937>.
- Chidambaranu, K., & Shunmuganathan, K. L. (2017). Predicting user preferences on changing trends and innovations using SVM based sentiment analysis. *Cluster Computing*, 1-5. <https://doi.org/10.1007/s10586-017-1505-0>.
- Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010, April). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World Wide Web* (pp. 661-670). ACM. <https://doi.org/10.1145/1772690.1772758>.
- Demirsoz, O., & Ozcan, R. (2017). Classification of news-related tweets. *Journal of Information Science*, 43(4), 509-524. <https://doi.org/10.1177/0165551516653082>.
- Ju, C., & Xu, C. (2013). A new collaborative recommendation approach based on users clustering using artificial bee colony algorithm. *The Scientific World Journal*, 2013. <https://doi.org/10.1155/2013/869658>.
- Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications*, 40(16), 6266-6282. <https://doi.org/10.1016/j.eswa.2013.05.057>.
- Chen, R. C., & Hsieh, C. H. (2006). Web page classification based on a support vector machine using a weighted vote schema. *Expert Systems with Applications*, 31(2), 427-435. <https://doi.org/10.1016/j.eswa.2005.09.079>.
- Wei, J., He, J., Chen, K., Zhou, Y., & Tang, Z. (2017). Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications*, 69, 29-39. <https://doi.org/10.1016/j.eswa.2016.09.040>.
- Zhang, K., Xin, X., Luo, P., & Guot, P. (2017, October). Fine-grained news recommendation by fusing matrix factorization, topic analysis and knowledge graph representation. In *Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on* (pp. 918-923). IEEE. <https://doi.org/10.1109/SMC.2017.8122727>.
- Viana, P., & Soares, M. M. (2017). A Hybrid Approach for Personalized News Recommendation in a Mobility Scenario Using Long-Short User Interest.
- Silva, J. B., Schreiber, J. N. C., & Nara, E. O. B. (2017). Bayesian approach to news recommendation systems. *Ciência da Informação*, 44(3).

- [18] Chen, C., Meng, X., Xu, Z., & Lukasiewicz, T. (2017). Location-Aware Personalized News Recommendation with Deep Semantic Analysis. *IEEE Access*, 5, 1624-1638. <https://doi.org/10.1109/ACCESS.2017.2655150>.
- [19] Haruna, K., Ismail, M. A., Damiasih, D., Sutopo, J., & Herawan, T. (2017). A collaborative approach for research paper recommender system. *PloS one*, 12(10), e0184516. <https://doi.org/10.1371/journal.pone.0184516>.
- [20] Meguebli, Y., Kacimi, M., Doan, B. L., & Popineau, F. (2017). Towards better news article recommendation. *World Wide Web*, 20(6), 1293-1312. <https://doi.org/10.1007/s11280-017-0436-2>.
- [21] Zhang, S., Yao, L., & Sun, A. (2017). Deep learning based recommender system: A survey and new perspectives. *arXiv preprint arXiv:1707.07435*.
- [22] Gupta, S., Sodhani, S., Patel, D., & Banerjee, B. (2017). Building a Dynamic News Category Network for News Sources Recommendations. *World Academy of Science, Engineering and Technology, International Journal of Computer and Information Engineering*, 4(7).
- [23] Wei, J., He, J., Chen, K., Zhou, Y., & Tang, Z. (2017). Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications*.
- [24] Portugal, I., Alencar, P., & Cowan, D. (2017). The use of machine learning algorithms in recommender systems: a systematic review. *Expert Systems with Applications*.
- [25] Jain, A., & Gupta, C. (2018). Fuzzy Logic in Recommender Systems. In *Fuzzy Logic Augmentation of Neural and Optimization Algorithms: Theoretical Aspects and Real Applications* (pp. 255-273). Springer, Cham. [https://doi.org/10.1007/978-3-319-71008-2\\_20](https://doi.org/10.1007/978-3-319-71008-2_20).