Deep Recurrent Network Based Feature Selection using Single Matrix Normalization and Eigen Vectors for Analyzing Sentiments

Rajalaxmi Hegde, Sandeepkumar Hegde

Abstract: ***Sentiment analysis plays a major role in e-commerce and social media these days. Due to the increasing growth of social media, a huge number of peoples and users send their reviews through the Internet and several other sources. Analyzing this data is challenging in today's life. In this paper new normalization based feature selection method is proposed and the topic of interest here is to select the relevant features and perform the classification of the data and find the accuracy. Stability of the data is considered as the most important challenge in analyzing the sentiments. In this paper investigating the sentiments and selecting the relevant features from the data set places a major role. The aim is to work with the vector-based feature selection and check the classification performance using recurrent networks. In this paper, text mining depends on feature retrieval methods to improve accuracy and propose a single matrix normalization method to reduce the dimensions. The proposed method performs data preprocessing or sentiment classification and features reduction to improve accuracy. The proposed method achieves better accuracy than the N-gram feature selection method. The experimental results show that the proposed method has better accuracy than other traditional feature selection approaches and that the proposed method can decrease the implementation time.***

***Index Terms*: *Features, Classifier, Sentiment, Normalization, Recurrent Networks, Covariance Matrix, Cosine similarity*.**

I. INTRODUCTION

These days a huge amount of data is obtained from websites customer blogs, reviews, etc. these product reviews play a major role in analyzing the sentiments of users. There are several methods to convert and analyze these structured and unstructured data, a lot of surveys are still going on to analyze the sentiments automatically from these reviews. Many researchers are gradually focusing on extracting the opinions and analyzing sentiments. Feature selection is considered as an important step in several machine learning algorithms, where selecting the relevant features improves the performance since it increases the performance of the model. Some other benefits in selecting the important features are to build the model and provide a better understanding. feature selection methods mainly consist of a filter and wrapper approach filter method directly apply on the dataset and provide feature weighting these methods are fast methods. Classification is done using cross-validation. The recurrent neural network is used in understanding the long term dependencies among the data. The Recurrent network is used to obtain the sequential relation among data. In the paper initially, the data is pre-processed and then data is vectorized using the proposed feature selection method and data is passed to the neural network model for further classification. In this, a method symmetric matrix normalization is proposed along with term document identification using weight matrices through their single matrix normalization. The proposed method includes two steps to obtain the word vectors and implementation of the semantic index and obtain the summarized result. Each review is tokenized and vocabulary of words is created. Feature vectors are being extracted depending o location of words present in vocabulary. The weighting scheme that is most often implemented is the term frequency – inverted document frequency (TFIDF). Each review is tokenized and vocabulary of words is created. Feature vectors are being extracted depending o location of words present in vocabulary. Proposed methods improve model performance and reduce computation cost.

Revised Manuscript Received on. August 02, 2019

Rajalaxmi Hegde, Department of CSE, NMAM Institute of Technology, Nitte, India.

Sandeepkumar Hegde, Department of CSE, NMAM Institute of Technology, Nitte, India.

II. II. LITERATURE REVIEW

Sentiment analysis is the field of Natural language processing plays and is one of the most important aspects of research these days. There is a huge increase in the number of product, blogs, forums. Several machine learning algorithms are proposed. [1,2] such as SVM NB and ME. Several combinations of unigram bigram and trigram combinations are proposed. SVM and NB classifier was proposed in the classification of the sentence. With accuracy 87.0. In [3] a new work is proposed to analyze reviews using CNET to perform classification. They use score methods to check the reviews are positive or negative. In [9] wordnet is used for extracting sentiments. Several classifiers such as SVM, NB, and DT is used to perform the classification of reviews. Proposed new approach [10] for sentiment classification using the fuzzy method. In [4] several preprocessing techniques such as tokenization, Pos tagging, and removal of stop words, applied on several machine learning algorithm NB, SVM, to obtain feature set[5.6]. And performed random subspace and bagging subspace technique. In [11] other techniques such as correlation chi-square, Information gain.

In [8] hybrid method is proposed for feature selection using SVM and NB classifier and k fold cross validation for classifying the sentiment of review documents. In [9] k means is proposed for clustering and then perform classification of data.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The feature selection stage refines features and provides accurate results. In this sentiment, classification is done using the long life method.[17] deep learning based document classification is proposed in [18]. Several information retrieval methods such as text classification and document clustering are proposed in [19]. Emotional features and the sentiment word embeddings are proposed. In [7] experiments are conducted by applying mutual information feature selection with Naïve Bayes (NB). There are several techniques of machine learning in sentiment analysis they are lexicon machine learning and rule-based method. [12] Ml methods use different machine learning algorithms and labeled data to train a classifier to find sentiment. In [13] this paper deep convolution with text normalization and character level embedding, for both unstructured and structured data is conducted to perform sentiment analysis to handle the less memory space.[14]In this method the word are classified using maximum entropy based classification using LSA and probabilistic method, where the similarity among the words are computed. In [15] a new sentiment analysis method is proposed using a hidden Markov model where the sentiment lexicon is used instead of a sequence of words. In this paper[16] text classification is done using Arabic documents. Chi-square method is used as feature selection to enhance the performance.

III. PROPOSED METHODOLOGY

The proposed method is shown in the figure1 where initial input reviews are taken and the data preprocessing stage represents the transformation of unstructured data into a structured collection. These represent the meaning of the document. The most basic method of the document represented as a vector of features obtained by preprocessing. In the proposed method novel word embedding technique is used which is based on weight and matrix analysis. This helps the syntactic structure index analysis algorithm to determine the summaries with good quality of model construction. Figure 1 shows the architecture of the proposed feature selection.

A. Data Pre-Processing and Feature Selection

The preprocessing stage represents the transformation of unstructured data into a structured collection. The most basic method of the document represented as a vector of features obtained by preprocessing. The weighting scheme that is most often implemented is the term frequency – inverted document frequency (TFIDF). It is calculated from the frequency of a feature f in document t d, the number of documents in the collection N and the number of documents, containing this feature df(f) as shown in the equation. The frequency mainly depends on a term in the document corpus helps in document discrimination. The more frequent the term appears in the documents in the corpus the lesser its contribution in document discrimination. The less frequent a term appears in the documents in the corpus the greater its Contribution in document discrimination. IDF i.e. inverse document frequency is a measure of the availability of a term in the document corpus. After applying the term weighting scheme, the matrix is decomposed using Single matrix normalization. Matrices obtained as a result of SMN are used to compute similarity scores between documents in the corpus. During performing preprocessing the text classification the number of Unique words in the given set is large and the vocabulary data set. Term weighting is used to obtain weights of individual terms in the documents. Term weight is a technique of identifying a respective document and in the given corpus. Figure 2 shows the proposed deep recurrent model.

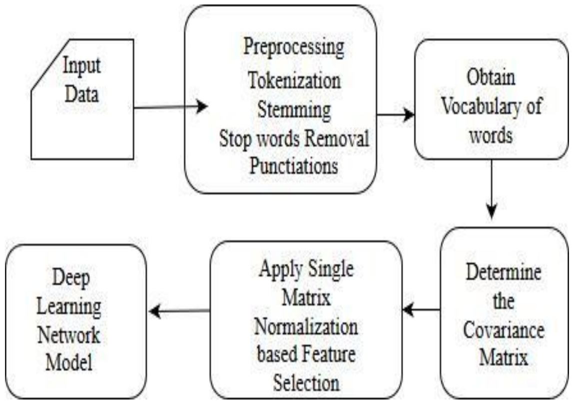

Fig. 1: Architecture of Proposed feature Selection

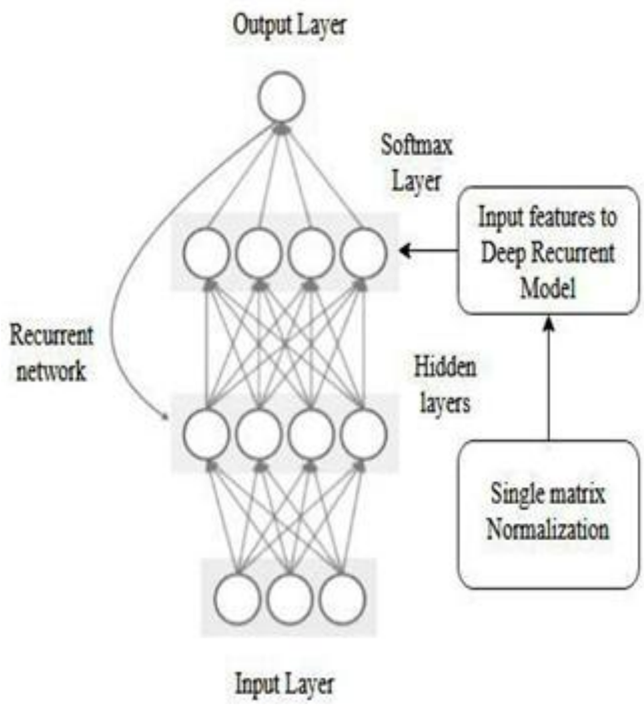

Fig 2: Proposed Deep Recurrent Network Model

Single Matrix normalization can either be used to retrieve all documents similar to a query or can be used to obtain percentage similarity between documents in a document corpus. As an initial step, the list of relevant words which is used for indexing is formed out of sample document A preprocessing is required to eliminate the irrelevant A Term-Document Matrix (TDM) is constructed with term frequencies of relevant terms. Documents may be of varying lengths ranging from very small to very large documents.

Long documents usually contain many occurrences of the same terms and occurrences of many different terms which increase their chance of retrieval. Term frequencies are hence normalized to reduce the impact of document length. There are different schemes for normalization. Another factor which is widely used for normalizing term frequencies is the maximum term count. Yet another scheme is cosine normalization. The normalization value of the features is computed using the below equation 1 and 2. The normalized length of document N*l* is calculated using equation 1. Where s denotes the document score before normalization and x denotes tuning parameter for normalization. The value of x usually in the range 0 to 1.

$$N_l = S \times \frac{1}{(1-x)+x\left(\frac{ld}{avld}\right)} \quad (1)$$

If the value of x is 1 then the equation is shown in 4.

$$N_l = S \times \left(\frac{avld}{ld}\right) \quad (2)$$

Let F_{wj} represents the feature weighs where weight updates are given by ΔFwj equation

$$\Delta F_{wj} = \delta \frac{\partial J}{\partial F_{wj}} = \delta \sum_i \left(y^{(i)} - z^{(i)}\right) F_j^{(i)} \quad (3)$$

$$F_{wj} = F_{wj} + \Delta F_{wj} \quad (4)$$

Where δ represents the learning rate which is the hyper parameter. y indicates the class label and Z denotes the actual output.

Our aim is to perform scaling of covariance on each pair of data by the product of deviations on each values. As shown in equation 4 and 5.

$$\alpha_{pq} = \frac{1}{r-1} \sum_i^r \left(p_i - \overline{p}\right)\left(q_i - \overline{q}\right) \quad (5)$$

Where α_{pq} indicates the covariance value before standardization.

$$\alpha'_{pq} = \frac{1}{r-1} \sum_i^r \left(\frac{p-\overline{p}}{\alpha_p}\right)\left(\frac{q-\overline{q}}{\alpha_q}\right) \quad (6)$$

$$\alpha'_{pq} = \frac{\alpha_{pq}}{\alpha_p \alpha_q} \quad (7)$$

Where α’pq indicates the deviation value after standardization.

The covariance matrix is denoted by M

$$M = \begin{pmatrix} \alpha(p,p) & \alpha(p,q) \\ \alpha(q,p) & \alpha(q,q) \end{pmatrix} \quad (8)$$

If p is positively correlated with q then q is also highly correlated with p. Let Rt represents the transformed matrix and Z indicates the scaled matrix.

$$R_t = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad (9)$$

$$Z = \begin{pmatrix} z_p & 0 \\ 0 & z_q \end{pmatrix} \quad (10)$$

Where θ represents the angle of rotation and Zp and Z_q represents the scaling factors.

$$M = \begin{pmatrix} \alpha_{p^2} & 0 \\ 0 & \alpha_{q^2} \end{pmatrix} \quad (11)$$

Indicates the covariance matrix in Eigen values as shown in equation 11.

$$M = \delta D \delta^{-1} \quad (12)$$

Where D denotes the diagonal matrix whose non zero elements will be the actual or corresponding Eigen values. M represents the Eigen decomposition of Covariance Matrix in equation 11.

Here δ indicates the matrix whose columns are Eigen vectors of M.

$$M = R_t Z Z R_t^{-1} \quad (13)$$

$$Z = \sqrt{D}$$

$$M = T_f T_f^{\,T} \quad (14)$$

Where M denotes the final covariance matrix obtained after several transformation and normalization.

The cosine similarity is computed using the equation 15.

$$cs = \cos\theta = \frac{p.q}{\|p\|\|q\|} = \frac{\sum_{i=1}^r p_i q_i}{\sqrt{\sum_{i=1}^r p_i^2}\sqrt{\sum_{i=1}^r q_i^2}} \quad (15)$$

The equation 15 denotes the computation of cosine similarity among the two corresponding values. Single matrix normalization is a process of analyzing relationships within documents and which is present by producing some contents related to documents.

The proposed method can be applied to any data over the discrete value so-called two-factor representation. Let D = {d1, . . . , dn} indicates the collection of documents with terms from W = {w1 …wm} We can build a matrix with dimension q x p and Mij represents the count of term wj present in document di. Di indicates a row vector, while Wj denotes column vector. The equation 15 denotes the computation of cosine similarity among the two corresponding values.

Pseudo Code to Compute Optimization using Single matrix Normalization

1. Let matrix M be the matrix of the given dimension p and q, an orthogonal matrix is o is obtained with dimension p and p where pxp and qxq are two orthogonal matrices and $p \times q$ diagonal matrix [Σ]

2. Obtain the Single matrix normalization of M i.e factorization of $M=O\Sigma V^T$, where O and V are unitary, and $\sum=diag(x1,\ldots,xr)$,$r=min(p,q)$, with $x1\geq\ldots\geq xr\geq 0$. Where xr indicates singular values, U indicates the single vectors of left and V denotes the single vector of right.

3. Let the set rank of A is r, the SMN of M can be simplified as $M=O_r\sum_r V_r^T$, where Or is a p-by-r matrix with O representing first r columns Vr is a q-by-r with first r columns of V are present and ∑ a r-by-r denote first r single value.

4. If r is substituted by k(k<r), we get a new matrix Mk as Mk=Ok∑kVkT

5. Return Mk is an optimal lower-rank approximation

In process SMN a matrix is divided into three different orthogonal matrix O, a matrix with diagonal Σ, and a transpose of a matrix orthogonal matrix T. The matrix is a Single value with representation of pxq matrix M and it is of the form M = OΣVT Here O is a p ×p orthogonal value, Σ is a p × q denotes the matrix with a rectangle with entries in the matrix with non-negative real numbers on the diagonal, and VT is a qxq orthogonal matrix. The p columns of O, called the singular matrix with left side representing vectors of M, called as vectors with Eigen values and the q represents the columns of V, called the singular matrix with right values, are eigenvectors of MTM. The diagonal entries Σii of Σ denotes the single values on M. An p × q matrix has its value (p, q) indicating different single values. Dimensionality reduction obtained by keeping only the single values which are above the threshold value The new values obtained are Eigenvectors MTM. The maximum Eigen values will be the same as the largest value of the Eigenvectors. The singular values are arranged in decreasing order of a matrix, both right and left singular vectors are also arranged accordingly. This simplifies the process of dimensionality reduction. If only r largest single value is retained, then only the first r left and right individual vectors are retained thereby reducing O to p × r matrix, Σ to an r × nonnegative matrix and VT to an r × q matrix.

IV. EXPERIMENTAL RESULTS

The experiments are conducted using reviews data set. The data set contains a list of reviews and its summary along with positive and negative values. Initially, Data preprocessing is done to remove unwanted data from the data sets. Tokenization, stop word removal Experiments are conducted to remove stop words and to perform stemming. Initially, several N-grams such as unigrams, bigrams, and trigrams feature selection techniques are conducted and the accuracy values are noted. Figure 3 shows a comparison of the accuracy values of several classifiers.

Table I. Accuracy of Classifiers with only N-grams

Model/Features	unigram	Bigram	Trigram
Logistic Regression	0.8933	0.925	0.9324
Bernoulli NB	0.8736	0.882	0.8700
Perceptron	0.8610	0.907	0.9192

Fig 3: Accuracy of N gram Features with Classifiers

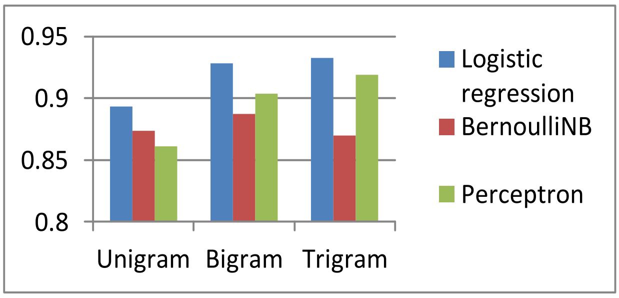

Table II. Classification Result of Positive values

	Precision	Recall	F-score
Logistic regression	0.76	0.42	0.54
Decision Tree	0.68	0.68	0.68
Perceptron	0.84	0.19	0.31
Bernoulli NB	0.88	0.07	0.13

Table III: Show classification Result Negativevalues

	Precision	Recall	F-score
Logistic regression	0.86	0.96	0.91
Decision Tree	0.91	0.91	0.91
Perceptron	0.81	0.99	0.98
Bernoulli NB	0.79	1.00	0.88

Table 1 denotes the Accuracy comparisons of N grams with several machine algorithms. Table 2 represents the precision, recall and F-score value of positive values. Table 3 represents the precision recall and accuracy of negative values.

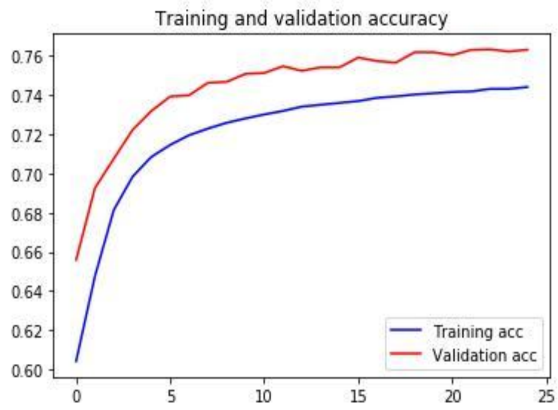

Fig 4: Training and Validation Accuracy of Deep Recurrent Model

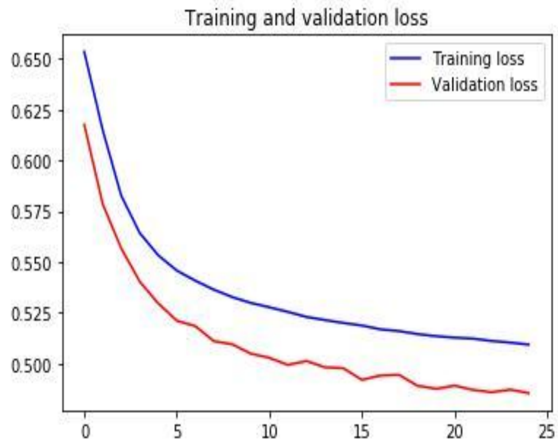

Fig 5: Training and validation loss of Deep Recurrent model

Figure 4 and Figure 5 represents the analysis curves of the proposed Deep recurrent network model. Figure 4 represents the training and validation accuracy which shows that validation accuracy is much better compared to training accuracy. Figure 5 shows that validation loss is getting gradually decreased as there is a decrease in the validation loss. Figure 6 shows the accuracy obtained for the proposed model which is 91.58% which is obtained by combining the Single matrix normalization based feature selection with the Deep neural network model.

```
Epoch 1/7  - 20s - loss: 0.4382 - acc: 0.8168
Epoch 2/7  - 19s - loss: 0.3214 - acc: 0.8648
Epoch 3/7  - 19s - loss: 0.2836 - acc: 0.8808
Epoch 4/7  - 19s - loss: 0.2578 - acc: 0.8958
Epoch 5/7  - 19s - loss: 0.2288 - acc: 0.9036
Epoch 6/7  - 19s - loss: 0.2103 - acc: 0.9158
```

Fig 6: Accuracy of the proposed model

The epochs represent the forward and backward pass for training examples. Batch size is same as the epochs.

V. CONCLUSION

Feature extraction in sentiment analysis is one of the most important aspects which has to be done accordingly. In this paper novel, a single matrix normalization based feature selection is proposed. Performance of the machine learning algorithms in combination with the proposed method is discussed. The experimental results shows that accuracy obtained for the proposed feature selection with Deep recurrent model outperforms better than the feature selection done using the individual N grams methods. 91.58% accuracy whereas the other machine learning algorithms performs around 89% accuracy. Further work indicates feature selection in combination with TF-IDF and other algorithms to perform ensemble methods.

REFERENCES

1. Catal, Cagatay, and Mehmet Nangir. "A sentiment classification model based on multiple classifiers." Applied Soft Computing 50 (2017): 135-141.
2. Boiy, Erik, et al. "Automatic Sentiment Analysis in On-line Text." ELPUB. 2007.
3. Annett, Michelle, and Grzegorz Kondrak. "A comparison of sentiment analysis techniques: Polarizing movie blogs." Conference of the Canadian Society for Computational Studies of Intelligence. Springer, Berlin, Heidelberg, 2008.
4. Ye, Qiang, Ziqiong Zhang, and Rob Law. "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches." Expert systems with applications 36.3 (2009): 6527-6535.
5. Manek, Asha S., et al. "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier." World wide web 20.2 (2017): 135-154.
6. Whitehead, Matthew, and Larry Yaeger. "Sentiment mining using ensemble classification models." Innovations and advances in computer sciences and engineering. Springer, Dordrecht, 2010. 509-514.
7. Narayanan, Vivek, Ishan Arora, and Arjun Bhatia. "Fast and accurate sentiment classification using an enhanced Naive Bayes model." International Conference on Intelligent Data Engineering and Automated Learning. Springer, Berlin, Heidelberg, 2013.
8. Agarwal, Basant, and Namita Mittal. "Sentiment classification using rough set based hybrid feature selection." Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. 2013.
9. Kolog, Emmanuel Awuni, Calkin Suero Montero, and Tapani Toivonen. "Using machine learning for sentiment and social influence analysis in the text." International Conference on Information Theoretic Security. Springer, Cham, 2018.

10. Mouthami, K., K. Nirmala Devi,
11. and V. Murali Bhaskaran. "Sentiment analysis and classification based on textual reviews." 2013 international conference on Information communication and embedded systems (ICICES). IEEE, 2013.
12. Adel, Aisha, Nazlia Omar, and Adel Al-Shabi. "A Comparative Study Of Combined Feature Selection Methods For Arabic Text Classification." JCS 10.11 (2014): 2232-2239.
13. Collomb, Anaıs, et al. "A study and comparison of sentiment analysis methods for reputation evaluation." Rapport de recherche RR-LIRIS-2014-002 (2014)
14. Arora, Monika, and Vineet Kansal. "Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis." Social Network Analysis and Mining 9.1 (2019): 12.
15. Xie, Xin, et al. "An improved algorithm for sentiment analysis based on maximum entropy." Soft Computing 23.2 (2019): 599-611.
16. Kang, Mangi, Jaelim Ahn, and Kichun Lee. "Opinion mining using ensemble text hidden Markov models for text classification." Expert Systems with Applications 94 (2018): 218-227.
17. Bahassine, Said, et al. "Feature selection using an improved Chi-square for Arabic text classification." Journal of King Saud University-Computer and Information Sciences (2018).
18. Zhiyuan Chen, Nianzu Ma and Bing Liu. (2018). Lifelong Learning for Sentiment Classification, p1-8..
19. Navonil Majumder, Soujanya Poria, Alexander Gelbukh and Erik Cambria . (2017). Deep Learning-Based Document Modeling for Personality Detection from Text.iEEE iNTElliGENT SYSTEmS, p74-79.
20. Mehdi Allahyari. (2017). A Brief Survey of Text Mining Classification, Clustering and Extraction Techniques.KDD Bigdas, p1-13.

AUTHORS PROFILE

Mrs. Rajalaxmi Hegde is working as Assistant Professor in the Department of Computer Science &Engg at NMAM Institute of Technology. She has over 7 years of Teaching experience. She completed her B.E degree in Information Science& Engg and Master Degree in Computer Science & Engg from Visvesvaraya Technological University, Belagavi. She is Currently pursuing Ph.D. from Visvesvaraya Technological University. She has published various research papers in National and International Journals and conferences. She is member of ISTE.

Mr. Sandeepkumar Hegde is working as Assistant Professor in the Department of Computer Science &Engg at NMAM Institute of Technology. He has over 9 years of Teaching experience. He also served as Assistant System Engineer at Tata Consultancy services. He completed his B.E degree in Information Science& Engg and Master Degree in Computer Science& Engg from Visvesvaraya Technological University, Belagavi. He is Currently pursuing Ph.D. from Visvesvaraya Technological University. He has published various research papers in National and International Journals and conferences..