A Neural Network Based Approach for Sentimental Analysis on Amazon Product Reviews

S. John Livingston, B. S. Tamil Selvi, M. Thabeetha, C. Pushpa Grena, Clementia Shiny Jenifer

Abstract: ***A Sentiment is an opinion or thoughts stimulated by human feelings. On the other hand, the growth of internet technology allows everyone to share their opinions on social-media or micro-blogs. That's how Sentiment Analysis has come into the picture in recent days. Mainly sentiment analysis contributes to the online products, political victory, film hits and celebrity dominations on social networks. This paper focalizes on product testimonials for online shopping websites. The online product reviews datasets are employed in this study which is taken from Amazon.com. From that dataset, the customer reviews have been analyzed through NLP and the corresponding sentiments have been scrutinized. Machine learning is a niche technology that is implemented in most applications nowadays. The sentiment analysis system is based on Machine Learning Prediction Analysis where Neural Networks are involved. Neural Networks is a powerful algorithm in machine learning techniques, which has got an architecture similar to the human brain system. MLP Neural Networks is a kind of Neural Network algorithm that is been used in this sentiment analysis. Neural Networks works on a larger dataset and sentiment analysis is more efficient compared to other machine learning algorithms such as the KNN algorithm, Naïve Bayes. Hence the Amazon products recommendation system is built with Neural Networks for effective performance.***

***Index Terms*: Multi-Layer perceptron(MLP), Natural Language Processing (NLP), Product Review, Sentimental Analysis.**

I. INTRODUCTION

A. Demand on sentiment analysis:

The opinion or sentiment on products has a huge utility which gives direct feedback to make decisions on those products. Before the days of the internet, many of us have asked suggestions to our parents, friends, and relatives. After the development of the modernization, people have raised their verdicts on each other. That's the way to social-media developments. With the opinion's analysis, the product gets a better shape in their performance. For example, consider a watch whose prime motive is time measuring. Eventually, after people's conclusions, loads of features were supplemented with watches such as music, calendar, alarm, notification panel etc., after which it has turned to become a smart watch that can do all fun and pieces of stuff. The smart watch got its peculiarities based on people's appraisals from product review analysis. This is one of the uses of sentiment analysis to advance productivity and consumer needs. All the reviews and remarks are received through a feedback system or social media NLP processing.

B. Natural Language Processing:

NLP is nothing but processed information of a language and is being used to do text mining of the reviews present in the dataset. From the review's content, meaningful words are only being extracted using stop words or stemmer words or using separate word representation methods that are convenient for several languages. By means of these techniques, it is eventually possible to easily extract the required words and thus serve to promote the prediction of the sentiment of a corresponding review text. In the event of sentiment analysis, the preprocessed words are being persuaded into vectors employing vector representation algorithms. These vector values are further trained and tested with machine learning algorithms. The block diagram of Natural Language Processing is depicted in the fig. 1.0.

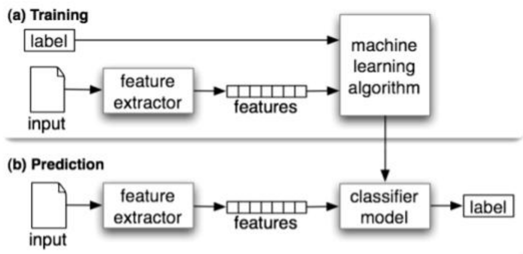


Fig. 1.0 Natural Language Processing

C. Machine Learning

Machine learning (ML) relies on patterns and inference instead of explicit instructions to efficiently execute a definite task by machine systems using the experimental study of algorithms and statistical figures. Machine learning calculations construct a scientific model of test information, known as "training data", to settle on expectations or choices without being expressly customized to

Revised Manuscript Received on April 11, 2019.

S. John Livingston, Department of Computer Science, Karunya Institute of Technology and Sciences, Coimbatore, India.
B. S. Tamil Selvi, Department of Computer Science, Karunya Institute of Technology and Sciences, Coimbatore, India.
M. Thabeetha, Department of Computer Science, Karunya Institute of Technology and Sciences, Coimbatore, India.
C. Pushpa Grena, Department of Computer Science, Karunya Institute of Technology and Sciences, Coimbatore, India.
Clementia Shiny Jenifer, Department of Computer Science, Karunya Institute of Technology and Sciences, Coimbatore, India.

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

perform out the task[1],[2]. Data mining is a field of concentrate inside machine learning and spotlights on exploratory information investigation through unsupervised learning[3],[4]. The various methods of machine learning algorithms differ in their methodology, the kind of information they yield, and the sort of query or task that they are relied upon to explain. It is of two types:

a) Supervised Learning: Supervised learning is the data represented with a piece of label information. For example, the cancer patient's symptom data has got two labels such as i) cancer affected ii) cancer not affected. With the label information, the symptom data are separated through a linear line or by means of other mathematical computation through which the machine can learn easily. If the dataset is large, the algorithms parameter should be tuned to fit the data otherwise the prediction method would produce the wrong output. Regression and Classification are included in Supervised learning techniques[5].When the outputs may include any numerical incentive inside a range, regression algorithms are being employed and when the outputs are bound to limited arrangement of qualities, classification algorithms are being employed.

b) Unsupervised Learning: Unsupervised learning algorithms focuses to discover structure in the data by only having the input values of a collection of data like clustering of data points and grouping. A huge amount of raw data is converted into clusters. Clustering is grouping the data by its similarity or properties. For example, K-means clustering and Fuzzy-clustering.

II. BACKGROUND AND LITERATURE REVIEW

One primary difficulty in sentiment analysis is the classification of sentiment polarity [6]-[10]. There is a difficulty to classify the text into one particular sentiment polarization, positive or negative (or neutral) of a given part of printed text. According to the data present, there exists almost three heights of sentiment polarity classification, to be specific, the sentence level, the document level, and the aspect and entity level [11]. The sentence level tackles with every sentence's sentiment categorization while the document level regards whether a document, expresses negative or positive sentiment as a whole.

Since audits of multiple tasks regarding sentiment analysis have already been included in [11], in this part, we will only examine some former work, upon which our analysis is primarily based. Hu paraphrased a record of positive words and Liu paraphrased a record of negative words, individually, based on consumer reviews[12]. The negative record holds 4783 terms and the positive record comprises 2006 terms. Both these records also do provide some misspelled terms that are usually present in social media and reviews of online products. Sentiment categorization is basically a classification difficulty in which the articles that hold opinions or sentiment message should be recognized before the analysis. For feature selection, Pang and Lee suggested eliminating objective sentences by extricating subjective ones[13]. They introduced a text-categorization procedure which is suitable to recognize subjective content using a minimal cut. Gann et al choose 6,799 labels based on Twitter data in which each token is specified a sentiment score, namely TSI(Total Sentiment Index), emphasizing itself as a positive label or a negative label [14]. Specifically, a TSI for a certain label is computed using "(1)":

$$TSI = \frac{p - tntp \times n}{p + tntp * n} \qquad (1)$$

where n is the number of times a label arises in negative tweets and p is the number of times a label arises in positive tweets. tntp is the ratio of the total number of negative tweets over the total number of positive tweets. The rating system for Amazon is described in the fig. 2.0.

Star Level	General Perception
★	I hate it.
★★	I don't Like it.
★★★	It's Okay.
★★★★	I Like it.
★★★★★	I Love it.

Fig. 2.0 Rating System for Amazon.com

III. EXISTING SYSTEM

A) Naïve Bayes Algorithm

NB algorithm runs based on the probability defined by the Bayes algorithm. The Bayes algorithm determines the review has how much probability to be positive and negative, with the observed probability NB algorithm picks the sentiment on the review comments. It works based on "(2)"

$$P(c \mid x) = P(x \mid c)P(c) / P(x) \qquad (2)$$

Where P(c|x) is the posterior probability of class (c, target) given predictor (x, attributes).P(c) is the prior probability of class(x|c) is the likelihood which is the probability of predictor given class(x) is the prior probability of predictor.

Henceforth NB robust algorithm in probability verdict has gained a low prediction accuracy over a large dataset. And NB blinds to the testing data which is not present in the training data. The probability outputs of NB have not attained much importance.

B) Support Vector Machine

A support-vector machine creates a hyperplane or set of hyperplanes in a huge or infinite-dimensional space, which can be utilized for classification, regression, or additional tasks like outliers' detection [15]. An example of support-vector hyperplane is portrayed in the figure 3.0. Intuitively, a good parting is accomplished by the hyperplane that has the greatest distance to the nearest training-data point of any class (so-called functional margin), since in imprecise the greater the margin, the lower the generalization error of the classifier [16].

Disadvantages:-

a) Demands much time on training for a large dataset.

b) Ineffective to noisy data.

c) Inefficient to Multi-label Classification.

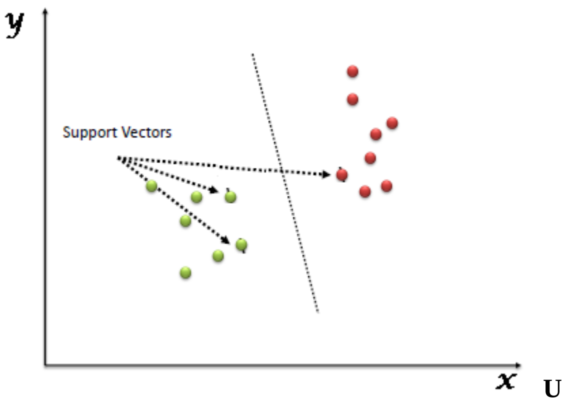


Fig. 3.0 Support Vector Hyper-plane

IV. PROPOSED SYSTEM

In this proposed system, the dataset is broken into Training data and Testing Data. Firstly, the review comments in the training data have been pre-processed, i.e., excluding by the stop words, stemming words, punctuation, special characters, and numeric data from reviews and at the end, only having the leftover meaningful words. Once, after cleaning the data, we extract feature vectors of the reviews. Count Vector and tf-idf vector algorithms are adopted in the computation of text to numeric data values. Count vector is a matrix of token counts which is transformed from a collection of documents. Term frequency-inverse document frequency is a numeric measure that is used to secure the value of a word in a record based on how frequently it appeared in that record and in the given collection of records. The intuition for this measure is: If a word appears repeatedly in a report, then it should be valuable, and we should mark that word with a high score. But if a word appears in too many other reports, it's presumably not a unique identifier; therefore, we should specify a lower score to that word. Refer "(3)" for the math formula for this test:

$$tfidf\left(t,d,D\right)=tf\left(t,d\right)\times idf\left(t,D\right) \quad (3)$$

Where t denotes the terms; d denotes each document; D denotes the collection of documents.

After building the feature vectors for each word into a matrix, it is exhibited as train vectors to the Multi-Layer Perceptron Neural network algorithm. The MLP algorithm is designed to fit the data size. Hence, corresponding to the feature vector quantity the number of hidden layers and neurons are assigned. Usually, the MLP algorithm feeds the feature vectors into input layers, and then it is transferred to hidden layers. In hidden layers, there are two steps that are exercised, one is applying weight to the vectors and the other is analyzing with test data. For applying weights to the data, MLP used the Adam Solver method which is also called as optimized stochastic gradient descent algorithm which computes weights for input feature vectors. Then Activation Function such as the Rectified Linear Unit (ReLu) compares the train vectors to the test vector and results in the class. And two more function MLP flow processes is Feed – forwarding & Back-propagation. In the input layer, the train vectors feed to the next layer i.e. hidden layers (forwards to all hidden layer one by one) then it processes the train vectors. Similarly, in the hidden layer, to reduce the error with respect to data size, the back-propagation method is applied to the train vectors to backpropagate again with re-assigned new weight inputs and thus minimizing the error rate. Fig.4.0 represents the classifiers comparison between true positive rate and false positive rate.

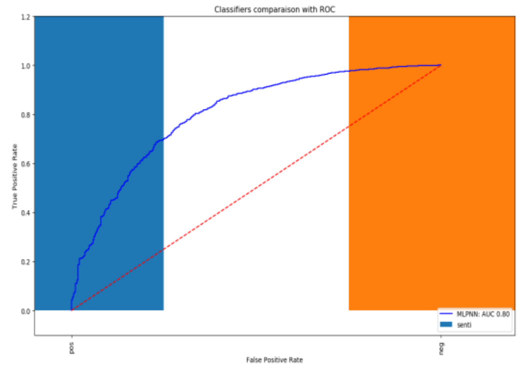


Fig. 4.0 Classifiers comparison

A) Data Preprocessing

The portrayal and nature of information is the essential worry before an analysis [17]. Usually, data preprocessing is the most significant aspect of a machine learning design, principally in computational biology [18].If there exists much extraneous and superfluous information present or in other words unreliable and noisy data, then the knowledge discovery through the training phase is more challenging. Data preparation and filtering steps can take a substantial quantity of processing time. Data preprocessing includes normalization, selection, feature extraction, cleaning, instance selection, and transformation, etc. The ultimate training set is nothing but the output of data preprocessing.

B) Feature Extraction

When a primary dataset is defined accurately and comprehensively while minimizing a collection of raw values to a much smaller group for processing, it is termed as Feature extraction which is a dimensionality compression method. The block diagram of feature extraction is depicted in fig. 4.1.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

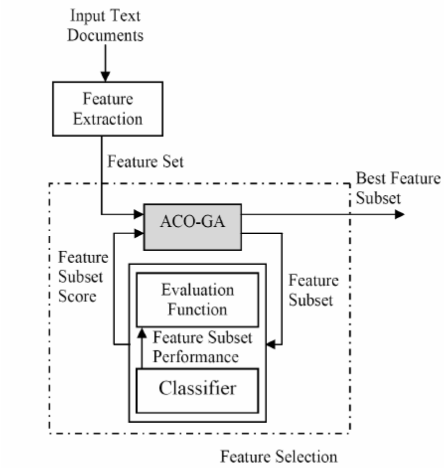


Fig. 4.1 Block diagram of feature extraction

It can be reconstructed into a condensed set of features (also called as a feature vector) when the algorithm's input data to is quite large for processing and it is contemplated to be tautological (example, the corresponding estimation meters and feet, or the repetitiveness of photographs displayed as pixels).Feature Selection is nothing but circumscribing a subset of the fundamental features[20]. For the desired task to be accomplished by using this condensed description in place of the comprehensive primary data, the selected features are demanded to accommodate the appropriate information from the input data.

C) Multi-Layer Perceptron

A multilayer perceptron (MLP) is a type of feedforward artificial neural network. Along with the input and output layers, Multilayer Perceptron (MLP) has also got one or more hidden layers. The structure of a multilayer perceptron is picturized in the fig.4.2. Using weighted links all these layers containing numerous neurons interconnected. The dataset's attributes are proportional to the number of neurons present in the input layer, the dataset's classes are proportional to the number of neurons present in the output layer. Besides the input nodes, a nonlinear activation function is practiced by every node which is a neuron.

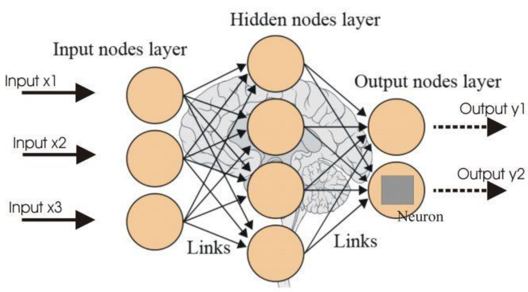
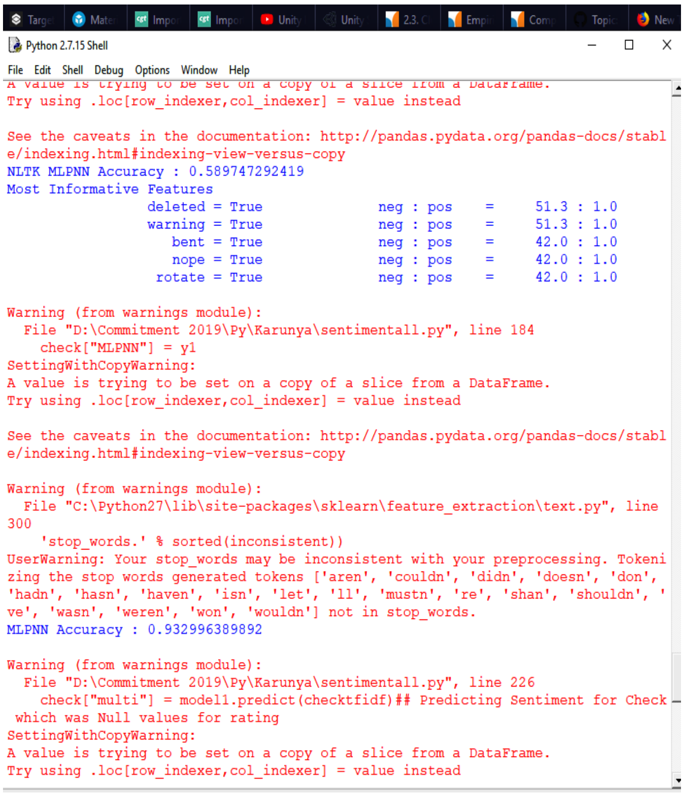


Fig. 4.2 Multi-layer Perceptron

Backpropagation for training is a superintended learning methodology employed by Multi-layer perceptron[21],[22]. A linear perceptron is distinguished by its multiple layers and non-linear activation. The non-linearly separable data can also be separated through this[23]. A "Vanilla" neural network is a name given to MLP at times, particularly when they possess a single hidden layer[24].

V.RESULTS AND ANNALYSIS

According to our dataset, the accuracy rate using Multi layer perceptron is analyzed to be 0.932996389892 which is more accurate and precise as compared to the Naïve Bayes and Support Vector Machine algorithms. The result of the accuracy rate is represented in fig.5.0.


```
Python 2.7.15 Shell
File  Edit  Shell  Debug  Options  Window  Help
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stabl
e/indexing.html#indexing-view-versus-copy
NLTK MLPNN Accuracy : 0.589747292419
Most Informative Features
                 deleted = True              neg : pos    =     51.3 : 1.0
                 warning = True              neg : pos    =     51.3 : 1.0
                    bent = True              neg : pos    =     42.0 : 1.0
                    nope = True              neg : pos    =     42.0 : 1.0
                  rotate = True              neg : pos    =     42.0 : 1.0

Warning (from warnings module):
  File "D:\Commitment 2019\Py\Karunya\sentimentall.py", line 184
    check["MLPNN"] = y1
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stabl
e/indexing.html#indexing-view-versus-copy

Warning (from warnings module):
  File "C:\Python27\lib\site-packages\sklearn\feature_extraction\text.py", line
300
    'stop_words.' % sorted(inconsistent))
UserWarning: Your stop_words may be inconsistent with your preprocessing. Tokeni
zing the stop words generated tokens ['aren', 'couldn', 'didn', 'doesn', 'don',
'hadn', 'hasn', 'haven', 'isn', 'let', 'll', 'mustn', 're', 'shan', 'shouldn', '
ve', 'wasn', 'weren', 'won', 'wouldn'] not in stop_words.
MLPNN Accuracy : 0.932996389892

Warning (from warnings module):
  File "D:\Commitment 2019\Py\Karunya\sentimentall.py", line 226
    check["multi"] = model1.predict(checktfidf)## Predicting Sentiment for Check
 which was Null values for rating
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

Fig. 5.0 Accuracy rate using MLP

REFERENCES

1. The definition "without being explicitly programmed" is often attributed to Arthur Samuel, who coined the term "machine learning" in 1959, but the phrase is not found verbatim in this publication, and may be a paraphrase that appeared later. Confer "Paraphrasing Arthur Samuel (1959), the question is: How can computers learn to solve problems without being explicitly programmed?" in *Koza, John R.; Bennett, Forrest H.; Andre, David; Keane, Martin A. (1996). Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. Artificial Intelligence in Design '96. Springer, Dordrecht. pp. 151–170. doi:10.1007/978-94-009-0279-4_9.*
2. *Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2*
3. Machine learning and pattern recognition "can be viewed as two facets of the same field."[2]:vii
4. *Friedman, Jerome H. (1998). "Data Mining and Statistics: What's the connection?". Computing Science and Statistics. **29** (1): 3–9.*
5. Alpaydin, Ethem (2010). *Introduction to Machine Learning*. MIT Press. p. 9. ISBN 978-0-262-01243-0.
6. Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr 2(1-2):1–135
7. Chesley P, Vincent B, Xu L, Srihari RK (2006) Using verbs and adjectives to automatically classify blog sentiment. Training 580(263):233

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

8. Choi Y, Cardie C (2009) Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2 - Volume 2, EMNLP '09. Association for Computational Linguistics, Stroudsburg, PA, USA. pp 590–598
9. Jiang L, Yu M, Zhou M, Liu X, Zhao T (2011) Target-dependent twitter sentiment classification. In: Proceedings of the 49th, Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, Stroudsburg, PA, USA. pp 151–160
10. Tan LK-W, Na J-C, Theng Y-L, Chang K (2011) Sentence-level sentiment polarity classification using a linguistic approach. In: Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation. Springer, Heidelberg, Germany. pp 77–87 Liu B (2012) Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers
11. Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, New York, NY, USA. pp 168–177
12. Pang B, Lee L (2004) A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics, ACL '04. Association for Computational Linguistics, Stroudsburg, PA, USA
13. Gann W-JK, Day J, Zhou S (2014) Twitter analytics for insider trading fraud detection system. In: Proceedings of the sencond ASE international conference on Big Data. ASE
14. *"1.4. Support Vector Machines — scikit-learn 0.20.2 documentation". Archived from the original on 2017-11-08. Retrieved 2017-11-08.*
15. Trevor Hastie, Robert Tibshirani, Jerome Friedman. "The elements of Statistical Learning", p. 134.
16. Pyle, D., 1999. *Data Preparation for Data Mining.* Morgan Kaufmann Publishers, Los Altos, California.
17. *Chicco D (December 2017). "Ten quick tips for machine learning in computational biology". BioData Mining. **10** (35): 1–17. doi:10.1186/s13040-017-0155-3. PMC 5721660. PMID 29234465*
18. *"What is Feature Extraction?". deepai.org.*
19. *Alpaydin, Ethem (2010). Introduction to Machine Learning. London: The MIT Press. p. 110. ISBN 978-0-262-01243-0. Retrieved 4 February 2017.*
20. Rosenblatt, Frank. x. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington DC, 1961
21. Rumelhart, David E., Geoffrey E. Hinton, and R. J. Williams. "Learning Internal Representations by Error Propagation". David E. Rumelhart, James L. McClelland, and the PDP research group. (editors), Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundation. MIT Press, 1986.
22. Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function *Mathematics of Control, Signals, and Systems*, 2(4), 303–314.
23. Hastie, Trevor. Tibshirani, Robert. Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, NY, 2009.

AUTHORS PROFILE

Mr. John Livingston.S, M.Tech.,(Ph.D.) is working as a Teaching faculty in department of Computer science and Engineering Karunya University, Coimbatore, South India. His area of research is Big data analytics, Image analysis and Interpretation. He is also an Associate member of IEI

Tamil Selvi is a final year Bachelor of Technology student in the department of Computer Science at Karunya Institute of Technology and Sciences, Coimbatore, South India.

Thabeetha is a final year Bachelor of Technology student in the department of Computer Science at Karunya Institute of Technology and Sciences, Coimbatore, South India.

C.Pushpa Grena is a final year Bachelor of Technology student in the department of Computer Science at Karunya Institute of Technology and Sciences, Coimbatore, South India.

Clementia Shiny Jenifer is a final year Bachelor of Technology student in the department of Computer Science at Karunya Institute of Technology and Sciences, Coimbatore, South India.