Volume 9, Issue 5
October 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

Reduction of Over-Fitting Problem in Predictive Forecasting Model using Deep Learning Neural Network

[1] Saritha K and [2]Dr. Sajimon Abraham

[1]School of Computer Sciences, Mahatma Gandhi University, Kottayam, Kerala, India

[2] School of Management and Business Studies, Mahatma Gandhi University, Kottayam, India
E-mail: [1]sarithakris@gmail.com, [2]sajimabraham@rediffmail.com

ABSTRACT

The large volume of complex and growing data generated from many distinct sources has led to the era of Big Data. In the case of Big Data, the commonly used software tools are not adequate to collect, manage and process the data within a tolerable elapsed time. Statistics as well as machine learning techniques are used for finding patterns and information from large data through data driven decision making. Big Data analytics gives competitive opportunities in designing business plans for Business Analytics. But the result of the analytics must be accurate and timely for successful decision making. The objective of this paper is to create a prediction forecasting model without having over-fitting problem on quantitative and categorical type data using Deep Learning Neural Network (DNN) technique. This paper creates two DNN network architectures with five layers including one input and one output layers. The first DNN model has one input layer, three hidden layers and one output layer. In the analysis of results obtained with this model shows that this model has over-fitting problem. The second DNN model is creates with similar network architecture. But this model includes the strategies for reducing the over-fitting problem with L2 regularization and Dropout methods. The data set collected from Kaggle repository are evaluated with the methodologies described in this work and we obtained more prediction accuracy with second DNN model.

Keywords: *Big Data Analytics, Deep Neural Network, Over-fitting, Dropout, Activation functions*

1. INTRODUCTION

The widespread use of digital technologies has led to the exponential growth of data from every imaginable source such as sensors, purchase transactions and social media networks. The large volume of complex and growing data generated from many distinct sources led to the era of Big Data. Companies depend on this massive data to take intelligence decisions as well as to gain a powerful competitive advantage. Modern society is also impacted by big data involving business, management, medical healthcare and government. Big data is extremely valuable to produce productivity in business and evolutionary breakthroughs in scientific disciplines, which give us a lot of opportunities to make great progress in many fields. In large volumes of data so much useful values are hidden and that can be generated only through the careful analysis. For this a new scientific paradigm has been born as Data Intensive Scientific Discovery (DISD) also known as Big Data Analytics.

Big Data analytics technologies and techniques should analyze the huge volume of data and generating conclusion from them to enhance the business and customer relationship [18]. The technology based analytical process for analyzing huge amount of data and presenting the information to help the end users to increase the decision making capacity. The data used in intelligent analysis is historical data as well as newly generated data from distinct sources. So the programs are the combination of following advanced analytic techniques: data mining, predictive analytics, statistical analysis, text mining, and Big Data Analytics.

This work discusses to create a deep learning algorithm for the predictive forecasting model. Deep learning, a subset of machine learning, utilizes a hierarchical level of artificial neural networks to carry out the process of machine learning. The artificial neural networks are built like the human brain, with neuron nodes connected together like a web. While traditional programs build analysis with data in a linear way, the hierarchical function of deep learning systems enables machines to process data with a nonlinear approach. In the present scenario the volume and the features of data is increasing. The selection of suitable features from the huge dimension of data set is a big problem and this is the main task in data preprocessing step in Neural Network (NN) model. Now the NN model is extended and the feature selection process is included as a part of the model itself [19].

The new architecture of the NN model is known as Deep Learning Neural Network (DNN) and the learning process is used with this model is known as deep learning algorithms. In the network architecture of DNN includes more than one hidden layers with many neurons. So this type of NN architecture can handle more features of the huge amount of unstructured data. The first objective of this work is to create DNN network architecture with five layers including input layer and output layer for building a prediction forecasting model with quantitative and qualitative type of structured data set. The second objective of this work is to reduce over-fitting problem with two different strategies such as L2 regularization and Dropout. We optimize the activation functions with ReLu

Volume 9, Issue 5
October 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

and Linear for the better accuracy of the model. Also optimize different Dropout rates and find that 30% Dropout rate gives better result than the first DNN architecture.

The rest of this paper is arranged as follows: Section 2 covers related study. Section 3 gives detailed explanations of the methodologies used. Section 4 describes experimental evaluation with the collected data set. Section 5 gives a detailed result analysis and then the conclusion.

2. RELATED WORKS

There have been several previous studies that use machine learning models to predict the forecasting. Schmidhuber [10] give a detailed overview on the subfield Deep Learning in Artificial Neural Network. The topology of simple NN and RNN are described. Different neural network models starts from 1960 to 2013, their working principles, learning algorithms of each network model etc are discuss in detail. Culkin et. al [2] describes in their work with brief overview of the ideas in AI, deep learning, and then proceed to show how these models may be mapped to a canonical problem in finance, that of option pricing. Palm et. al [3] implement and evaluate state-of-the-art deep learning models and using these as building blocks for investigate the hypothesis that predicting the time-to-time sensory input is a good learning objective. They also discuss the computational power of Deep learning over simple machine learning model.

Sze V et al [13] present a brief chronology of the major steps in its history, and some current domains to which it is being applied. They also give a detailed study on the use of dropout in hidden layers. Zhang et al. [17] specifies Dropout optimization is very necessary for building the DNN model. Dropouts can dramatically improve the convergence and also reduce the testing error. The researchers recommended that the number of dropouts is around 20%-50% and used a 50% dropout as a baseline configuration of CCN's parameters. Dalto M, et al. [15], that the DNN model is more effective than NN model for classification problems as well as for regression problems.

3. METHODOLOGIES

Deep learning algorithms are machine learning methods based on distributed representations. Deep learning attempts to learn high-level features in data by using structures composed of multiple non-linear transformations. Deep Neural Network (DNN) model is the development of the NN where the number of layers in the hidden layer is more than one. DNN model has been widely used to solve problems related to speech recognition [14], pattern recognition, image recognition, and several other DNN applications that continue to be developed. Besides that, DNN is capable of learning high-level features with more complexity and abstract than shallow neural network [7].

In general, the architecture of the DNN in this study, adopts the general architecture, where there is a feature extraction layer that uses interchangeably between the hidden layers, followed by a classification or regression layer such as Multi-Layer Perceptron is based on backpropagation [4]. DNN model that examined in this research consists of two architectures. The proposed algorithm for implementing DNN architecture with handling over-fitting problem is described in Algorithm 1.

3.1 Preprocessing the Data Set

Preprocessing of the collected data set is an important part of supervised data analytics. Preprocessing of collected data set is very much required as it is difficult to produce accurate results with the noisy data. Different methodologies are required for this preprocessing.

- **Handling Missing Values**

There are several handle missing methods can use to overcome the problem. In this work we use to replace missing values with most frequent data. This method is an efficient on both quantitative and qualitative type of data set.

- **Normalization**

In this study, the second step of pre-processing data is normalizing data. Normalization is the process of scaling data to have a norm unit. Each sample (i.e., each row of the data matrix) with at least one nonzero component is transformed independently from the other sample and causes the norm (l1 or l2) is equal to one (the norm parameter l1 / l2 is used to normalize each nonzero sample). Then, we use a the min-max scalar with Equation 1, which scales the dataset so that all the input features lie between 0 and 1 inclusive.

$$y'_i = \frac{Y_{max} - Y_i}{Y_{max} - Y_{min}} \qquad (1)$$

- **Splitting the Data Set**

The third step of pre-processing data is splitting data. It has done with the aim of dividing the data into training and testing data with the proportion of 70%: 15%:15% as training data, validation data and test data.

3.2 Prediction Model with DNN

An artificial neural network is typically specified with An architecture: the width and depth of network, A learning algorithm: for updating the weight to model a task correctly and An activation unit: to transform a neuron's weighted input to its output activation. The same architecture is used in all experiments as discussed below. All the activation functions and learning algorithms used in this work have been formally introduced below.

- **An Architecture**

In this work we create the architecture of DNN in Figure1, consists of five layers including the input layer, three hidden layers and one output layer with one neuron. Two hidden layers that consist of the dense layer with 32 neurons.

Volume 9, Issue 5
October 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

The input layer of the DNN architecture is suitable with n dimensions. In the first hidden layers, it is consist of 32 neuron setting from the combination of n features. Also, the second and the third hidden layer are the dropout. This dropout layer aims to avoid over-fitting or under-fitting during model training. Here the output layer only consists of one neuron. This is because the problem to be solved is a regression.

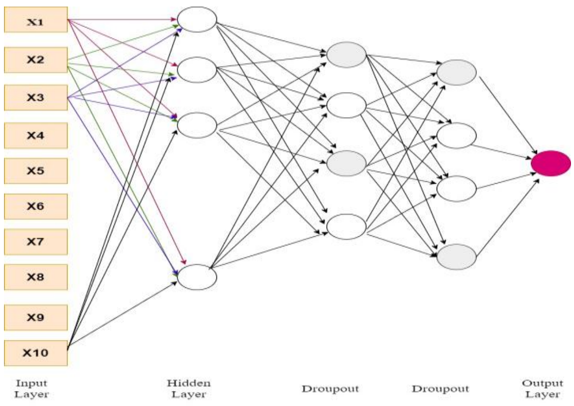


Figure 1 Architecture of DNN with three hidden layers

- **A Learning Algorithm**

Stochastic Gradient Descent is probably the most used learning algorithm for machine learning and particularly for deep learning [10]. The standard gradient descent algorithms update the parameters approximated by evaluating the cost and gradient over the full training set. The GD updates and computes the gradient of parameters using single or few training examples. The parameter update is given by:

$$\Delta W_{ij}\,(m)= \mu\delta_k\, S_j \qquad (2)$$

Where δ is the learning rate parameter and S is the input from hidden unit to output unit. The benefit of updating the parameters based on few training examples is, it reduces the variance in the parameter update and leads to a stable convergence. The crucial parameter of GD is learning rate which must be adjusted with a lot of trial and error. In this study, we have used default learning rate which is 0.01. The gradient descent algorithm is a method to find the weights for a multilayer feed forward neural network.

- **Activation Function**

The activation function is a function that making the layer active and mapping neurons from the input layer to neurons in the output layer. Here we use the activation functions as:

The Rectified Linear Unit (ReLu) activation function is defined by:

$$f(x) = \max(0,x) \qquad (3)$$

Where x is input to neurons. The output layer uses the linear activation function with equation.

$$Y=f(X)=X \qquad (4)$$

Here the output is same as the input and the function is defined in the range (- α, + α).

3.3 Handling Over-fitting

On the DNN model, the number of hidden layers is more than one and it certainly allows DNN to use a large number of parameters and tends to exhibit a highly complex cost function. Besides that, it certainly causes overfitting. If the prediction model has fit so extremely to the training data that it fails to generalize to other examples. The problem here was that training loss failed to give us an accurate picture of how well the model generalizes to unseen data. Due to the over-fitting problem the model that performs best on the training set, rather than the validation set model doesn't generalize well from our training data to unseen data.

In this work are combining two strategies of reducing over-fitting problem such as L2 regularization and Dropout methods. Dropout is a technique that randomly excludes neurons during the training process. The goal is to reduce the possibility of noisy neurons. L2 regularization is performed with the equation 5.

$$L2=\frac{\sum_{i=1}^{n}\left(y_i-\sum_{j=1}^{p}x_{ij}\beta_j\right)}{\textbf{Loss}}^{2}+\lambda\sum_{j=1}^{p}\boldsymbol{\beta_j}^{2} \qquad (5)$$

3.4 Accuracy Evaluation

Various accuracy evaluation methods are used for finding the accuracy of the proposed prediction model with training and testing data set. The difference between actual and predicted value is obtained with the equation 6.

$$MSE = 1/n\sum (Y_i - \hat{Y}_i)^2 \qquad (6)$$

Goodness of fit of the mode with accuracy is obtained with the equation R^2

$$R^2 = \frac{\sum_{i=1}^{n}(Y_i-\bar{Y_i})^2-\sum_{i=1}^{n}(Y_i-\hat{Y_i})^2}{\sum_{i=1}^{n}(Y_i-\bar{Y_i})^2} \qquad (7)$$

4. EXPERIMENTAL EVALUATION

For implementing the above proposed methodology for prediction model with the dataset available in Zillow's Home Value Prediction Kaggle competition data [5].

4.1 Preprocessing the data set

The number of attributes and tuples were 11 and 17379 respectively. In this study, three steps of data processing were carried out. We want to extract out the first 10 columns as features x1,….x10 and the label of what we want to predict as Y.

Volume 9, Issue 5
October 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

Algorithm 1: DNN-Predict (Deep Neural Network for prediction)

Input:

The Input Data as DF ← *Read the data set with nXm dimension*

Output:

Proposed Prediction model DNN-P

Method:

1. ***Preprocess the data set for handling missing values***
   // Normalize the data set
2. ***For i=1 to n do***
   $$Y'_i = \frac{Y_{max} - Y_i}{Y_{max} - Y_{min}}$$
   End For
3. ***Split the data set as X_Train, X_Val, X,Test, Y_Train, Y_Val and Y_Test***
   // Create the baseline sequential DNN model
4. ***Create the first input layer with n neurons***
5. ***Create the first dense hidden layer with (n*3)+2 neurons***
6. ***Create the second dense hidden layer with (n*3)+2 neurons***
7. ***Create the third dense hidden layer with ((n*3)+2)/8 neurons***
8. ***Create the last output layer with 1 neuron***
9. ***For i=1 to n, j=1 to m do***
   $$Y_{ij} = W_{ij} X_{ij} + b$$
10. ***End For***
    // Perform weight updating
11. ***For i=1 to n, j=1 to m do***
    $$W_{ij}(m+1) = W_{ij}(m) + \mu \mathcal{E}_j a_i$$
    $Y_{ij} = max(Y_{ij}, 0)$; *// Rectified Linear Unit*
12. ***End For***

- **Handling missing values**

fillna is a function in Python for handling missing value problem. In this work the missing values are replaced with most-frequent data.

- **Normalization**

Then, we use the min-max scalar with equation 1, which scales the dataset so that all the input features lie between 0 and 1 inclusive.

- **Splitting ratio**

The experiment is conducted on dataset. The experiments depicts that the more the trained data, more accurate the prediction accuracy. So the Training:Testing:Validation split ratio 75:15:15 shows better accuracy values and so this split size is considered for the remaining experiments.

4.2 Model Building

This work deals with the creation of two architecture of DNN model for prediction. These architectures are implemented through Python with Keras sequential model. Compare the results obtained with two DNN architecture and find that the proposed DNN prediction model shows more accuracy.

- **First DNN Model**

The architecture of this DNN is similar to explained in methodology part with one input layer, three hidden layers and one output layer. The hidden layers use the activation functions with equations 3 and 4.

- **Second DNN model**

The proposed second architecture is similar to the first model. But the strategies for reducing the over-fitting problem explained in the methodology part are also incorporated with the hidden layers. Different probability of excluding the neurons in the hidden layers such as 20%, 30% and 400% are evaluated. From these evaluations we got 30% dropout gives more accuracy.

5. RESULT ANALYSIS

In this study, several processes of DNN is carried out. The first step is the process of optimizing the DNN architecture model respectively. After the model of the DNN is optimized, the second step is comparing the two DNN and NN model based on the loss value of models to predict the actual value.

5.1 Optimization of DNN model

Different parameters explained in the methodology part are optimized with two DNN architectures. Some optimization processes for the DNN model are explained as follows.

- **Visualize loss and accuracy**

Figure 2 shows the graphical representation of loss and accuracy of the first DNN architecture prediction model. Training loss, validation loss and test loss are shows in Figure 2(a) and 2(b). Model accuracy on training, validation and test data set shows in Figure 2(c) and 2(d). These graphs show the clear sign of over-fitting. The training loss is decreasing, but the validation loss is way above the training loss and increasing. Also there is huge difference between training and test loss. We can see a clearer divergence between train and validation accuracy. The divergence of accuracy with training and testing is very high and it is seen in the Figure 2(d). So we can conclude that there is an over-fitting problem.

5.2 Optimization of DNN Model without Over-fitting problem

L2 regularization and 30% Dropout of neurons in the hidden layers are performed with second DNN architecture. The optimization process for this DNN model is explained as follows

- **Visualize loss and accuracy without over-fitting**

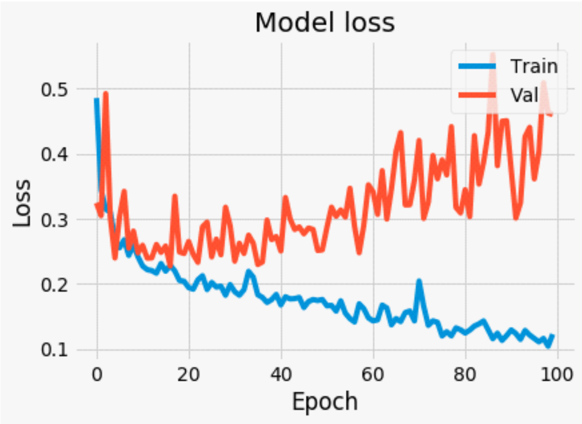
Figure 3 shows the graphical representation of loss and accuracy of the second DNN architecture. Training loss and validation loss are shows in Figure 3(a). Model accuracy on training and validation set shows in Figure 3(b).

From Figure 3(a) it is clear that the validation loss much more closely matches the training loss with 100 epochs. We can see accuracy between train and validation set data with the graph in Figure 3(b). Compare the two models; DNN Model 2 is reducing over-fitting substantially.

Volume 9, Issue 5
October 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

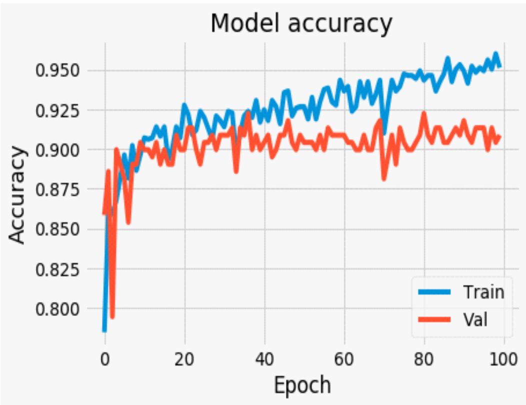


(a)

Figure 2(a) Plot for model loss with Training and Validation set data

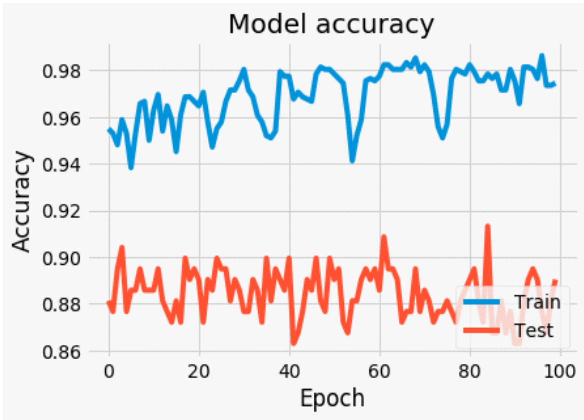
(b)

Figure 2(b) Plot for model loss with Training and Test data



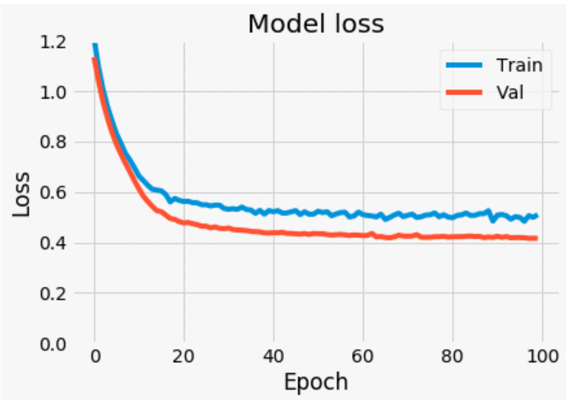
(c)

Figure 2 (c) Plot for model accuracy with Training and Validation set data



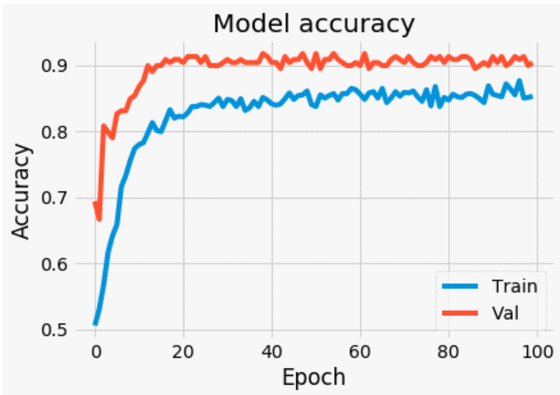
(d)

Figure 2(d) Plot for model accuracy with Training and Test data



(a)

Figure 3. (a) Plot for model loss



(b)

Figure 3. (b) Plot for model accuracy

Volume 9, Issue 5
October 2020

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

5.3 Performance Comparison with 2 DNN Models with NN Model

In the performance comparison part of the work is obtained with the comparison of prediction accuracy with the two DNN models and the simple NN model. In the NN model, we use SGD (0.01) as optimizer, one intermediary layer with 32 and ReLu activation function and one output neuron with linear activation function. This model is Multilayer Perceptron Neural Network with gradient Descent learning (MPNN-GD) [7].

Table 1 shows the statistical accuracy measures R^2 values calculated with equation 7. The graphical representation of the accuracy measures are show in the Figure 4.

Table 1. Statistical Accuracy measures

Modal Name	Train Accuracy	Validation Accuracy	Test Accuracy
MPNN-GD	88.1	70.62	92.01
Model 1 with DNN	90.87	70.62	62.2
Model 2 with L2 and Dropout	89.95	90.41	90.01

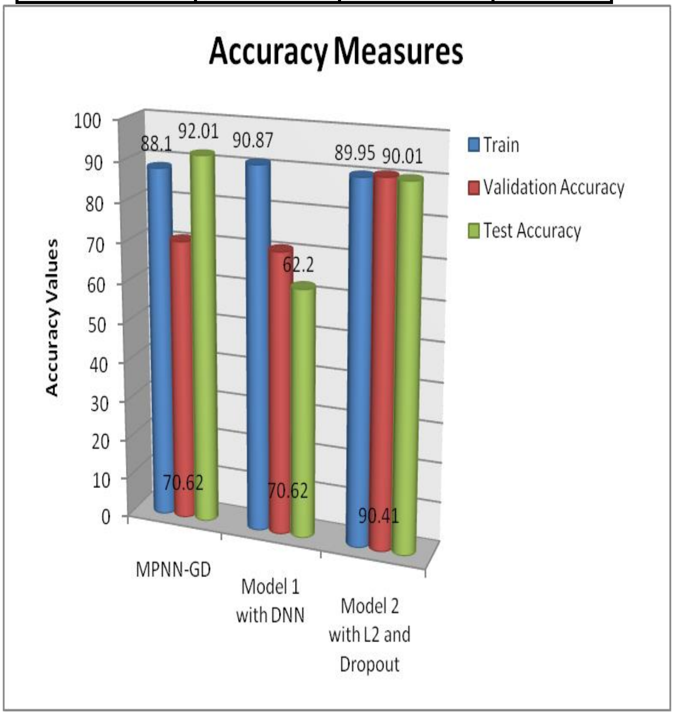


Figure 4. Comparison of Accuracy Measures

Based on the analysis of the accuracy value in table 1, it can see that the second DNN architecture gives better accuracy value compared with the first DNN architecture and the NN model. It means that the more dropout layers on the DNN architecture, the better the performance of DNN model. We guess that it caused by the ability of parameter dropout to reduce the possibility of noisy neurons, help prevent over-fitting and increase the performance of DNN model. The second DNN model without over-fitting problem have the accuracy measures with training, testing and validation set is very closer compared to other two.

6. CONCLUSION

For quantitative data analysis on structured data set is normally performed with statistical linear regression techniques. But in the case of Big Data the huge size of quantitative and qualitative type of structured data set will not give better accuracy of prediction results with this technique. So in this work create a DNN machine learning network architecture for creating a prediction forecasting model with such type of data set. The first architecture of DNN model has over-fitting problems. The second architecture of the DNN model that used with a 30% dropout rate in each of second and third hidden layers and also perform L2 regularization for reducing over-fitting problem. Compare these two results with simple Neural Network model. The optimized DNN architecture with dropout and regularization produces a smaller loss value compared with NN.

REFERENCES

[1] Wu, Xindong, et al. "Data mining with big data." IEEE transactions on knowledge and data engineering 26.1 (2013): 97-107.

[2] Culkin, Robert, and Sanjiv R. Das. "Machine learning in finance: the case of deep learning for option pricing." Journal of Investment Management 15.4 (2017): 92-100.

[3] Palm, Rasmus Berg. "Prediction as a candidate for learning deep hierarchical models of data." Technical University of Denmark 5 (2012).

[4] Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." Neural networks 61 (2015): 85-117.

[5] https://www.kaggle.com/c/zillow-prize-1/data

[6] C.L.Philip Chen,Chun-Yang Zhang, Data Intensive applications, challenges, and technologies: a survey on Big Data, Information Sciences 275 (2014) 314-347.

[7] Saritha, K., & Abraham, S, "Accuracy Evaluation of prediction using supervised Learning Techniques", Proceeding of the third international conference on advanced informatics for computing research, 2019, ACM- ISBN- 978-14503-6652-6

[8] Saritha, K., & Abraham, S. "Big Data Challenges and

Issues: Review on Analytic Techniques”. Indian Journal of Computer Science and Engineering (IJCSE) Vol. 8 No. 3 Jun-Jul 2017, p-ISSN : 2231-3850, e- ISSN: 0976-5166, (UGC Approved).

[9] Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 Dropout: A Simple Way to Prevent Neural Networks from Overfitting The Journal of Machine Learning Research 15 pp 1929-1958.

[10] Dutt, Micheal, Vimala Nunavath, and Morten Goodwin. "A Multi-layer Feed Forward Neural Network Approach for Diagnosing Diabetes." 2018 11th International Conference on Developments in eSystems Engineering (DeSE). IEEE, 2018.

[11] Mohaimenuzzaman Md, Abdallah Z S, Kamruzzaman J and Srinivasan B 2018 Effect of HyperParameter Optimization on the Deep Learning Model Propose for Distributed Attack Detection in Internet of Things Environment [arXiv:1806.07057]

[12] Li Z, Gong B and Yang T 2016 Improved Dropout for Shallow and Deep Learning NIPS’16 Proceeding 30th Conference on Neural Information Processing Systems pp 2531-2539

[13] Sze V, Chen Y, Yang T and Emer J S 2017 Efficient Processing of Deep Neural Network: A Tutorial and Survey Proceeding of the IEEE 105 pp 2295-2329.

[14] Yin S, Liu C, Lin Y, Wang D, Tejedor J, Zheng T F and Li Y 2015 Noisy Training For Deep Neural Network In Speech Recognition EURASIP Journal on Audio, Speech, and Music Processing 2 pp 1-14

[15] Dalto M, Jadranko M AND Vasak M 2015 Deep Neural Network for Ultra Short term wind forecasting Proceeding of IEEE pp 1657-1663

[16] Bianchini M and Scarselli F 2014 On the Complexity of shallow and deep neural network classifier: A comparison between shallow and deep architectures IEEE Transactions on Neural Network and Learning Systems 25 pp 1553-1565

[17] Zhang Y and Wallace B C 2016 A Sensitivity Analysis Of (Practitioners’ Guide To) Convolution Neural Network For Sentence Classification Proceedings of the 8th International Joint Conference on Natural Language Processing pp 253-263

[18] Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. " O'Reilly Media, Inc.".

[19] Isabelle G and Elisseeff 2003 An Introduction to Variable and Feature Selection Journal of Machine Learning Research 3 pp 1157-82

AUTHORS PROFILE

Mr Sajimon Abraham. MCA, MSc(Mathematics), MBA, PhD (Computer Science). He has been working as Faculty Member in Computer Applications & IT, School of Management and Business Studies, Mahatma Gandhi University, Kottayam, Kerala, India and also research guide in Mahatma Gandhi University. He currently holds the additional charge of Director (Hon), University Center for International Cooperation. He was previously working as Systems Analyst in Institute of Human Resource Development, Faculty member of Computer Applications in Marian College, Kuttikkanam, Kerala, India and Database Architect in Royal University of Bhutan under Colombo Plan on deputation through Ministry of External Affairs, Govt of India. His research area includes Data Science, SpatioTemporal Databases, Mobility Mining, Sentiment Analysis, Big Data Analytics and E-learning and has published 60 articles in National, International Journals and Conference Proceedings.

Ms Saritha K pursed Bachelor of Science from Mahatma Gandhi Universty, Kottayam, Kerala, Master of Computer Application from Bharathidasan University, Trichy, Thamilnadu and M.Phil (Computer Science) from Bharathiyar University, Tamilnadu. She is currently pursuing Ph.D. and working as Faculty Member in Department of Computer Science, School of Technology and Applied Sciences, Kottayam, Kerala. She has 17 years of teaching experience and two years of research experience. Her main research area focuses on Data Science, Data Mining Statistical Data Analytics and Neural Networks.