

## COMPARATIVE ANALYSIS OF CLASSIFIERS FOR EDUCATION CASE STUDY

Nurshahirah Abdul Malik\*<sup>1</sup>, Mohd Shahizan Othman<sup>2</sup> and Lizawati Mi Yusuf<sup>3</sup>

<sup>1, 2, 3</sup> Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia.

shahirahmalik@gmail.com, shahizan@utm.my, lizawati@utm.my

### ABSTRACT

Recently, classification is becoming a very valuable tool where a large amount of data is used on a wide range of decisions for the education sector. Classification is a method that used to group data based on predetermined characteristics. It is utilized to classify the item as indicated by the features for the predefined set of classes. The main significance of classification is to classify data from large datasets to find patterns out of it. Nevertheless, it is very important to choose the best classification algorithm which is also called as the classifier. Therefore, this research aims to conduct comparative evaluation between four classifiers which are Deep Neural Network (DNN), Random Forest (RF), Support Vector Machine (SVM) and Decision Tree (DT). All these classifiers have its own efficiency and have an important role in identifying the set of populations based on the training datasets. To choose the best classifiers among the four classifiers, the classifiers performance is required to be evaluated based on the performance metrics. The performance metrics of these classifiers were determined using accuracy and sensitivity rates. This study used education case study on student's performance data for two subjects, Mathematics and Portuguese from two Portugal secondary schools and data on the student's knowledge of Electrical DC Machines subject. After comparing the accuracy and sensitivity rates, DNN has the highest accuracy and sensitivity rate of classification and can be used to further the education-based research in future.

**Keywords:** Classifiers; Student Performance; Deep Neural Network; Support Vector Machine.

### I. INTRODUCTION

Toward the objective of creating a base-line is by carrying out the comparison in classification, this scenario will provide the information for appropriate model selection (Tan et al., 2006). Classification plays the important role in grouping the data based on specific criteria. The definition of classification is grouping data based on its format, type of analysis, processing techniques, and the data sources were required to retrieve, load, process, analyze and store (Suthaharan, 2014). As per Khasanah and Harwati (2017), classification as the undertaking of allocating articles to one of a few

predefined classifications that uses a lot of pre-characterized guide to manufacture a model that can order the enormous record population.

One good tip for choosing classification techniques is based on the applications being used and actual classification can be carried out with required data need to be obtained from a huge amount of data. There is no best classifier for all issues. The accuracy of the classifier changes dependent on the informational collection. Relationship between the indicator factors and the result is a key influencer. The decision should be made dependent on experimentation. Basically, the two main types of classification techniques are supervised and unsupervised classification. Supervised classification can be classified as predictive or directed classification. Meanwhile, unsupervised classification are descriptive or undirected whereby the type and number of the possible class are unknown, however, it able to assign a name to the class after carrying out classification.

This research sets its aim on conducting comparative evaluation between four classifiers based on their performance metrics. These four classifiers, Deep Neural Network (DNN), Random Forest (RF), Support Vector Machine (SVM) and Decision Tree (DT) are utilized in this research. Every classifier has its own efficiency and assists in recognizing the set of populations by referring the training dataset. The end goal for this research is to pick the best classifiers among them, the classifiers performance should be assessed based on the performance metrics.

This paper consists of six sections. Section I briefly explains the concept of the paper and related work is highlighted in section II. Section III presents the classifiers techniques. In section IV, the evaluation of methodology performance is discussed. Section V examines the experimental and results. The conclusion is finally made in Section VI.

## II. RELATED WORKS

In (Caruana & Niculescu-mizil, 2006), the utilization of an assortment of performance criteria to evaluate the learning techniques was observed. Ten supervised learning was assessed using eight performance criteria. It assess the execution of neural nets, logistic regression, SVM, Naive Bayes, memory-based learning, boosted trees, boosted stumps, bagged trees, random forests, and decision trees on eleven binary classification issues utilizing multiple performance metrics such as accuracy, cross-entropy, ROC area, lift, average precision, squared error, precision or recall break-even point, and F-score. Out of all the algorithms being used in this research, DT gives a better result to predict fertility index.

Ali et al. (2012) explain the difference in classification results between two model which is RF and DT to classify twenty versatile datasets. The classification parameters are made up of correctly and incorrectly classified instances, accuracy, F-Measure, recall and precision for 20 datasets that containing instances varying from 148 to 20000. It explain the advantages and disadvantages of using RF and DT for both small and large data sets. RF shows better performance for the larger data sets with a greater number of instances and number of attributes, while DT is useful with a lower number of instances with small data sets. The classification outputs present that RF is successfully used in the medical sector. Thus, this research wants to prove that RF is also can be used as a classifier in the education sector.

In this paper (Bhuyar, 2014), a classification algorithm that is different from the experiment was used in soil data set to predict its fertility. This paper discusses on the classification of soil fertility rate utilizing RF, DT, and Naïve Bayes algorithm. The

proposed work examines the use of classifiers and these calculations based on fertility index. Those data can be used for learning are valuable for decision makers and farmers to make the best choice. Evaluation criteria are based on time to build a model, effectively and inaccurately classified instances and prediction accuracy. It is shown that DT classifier performs better in predicting fertility index as the outcome of the research.

Although every classification method has its own edge, real-world issues failed fulfil the assumption of that specific method which is the approach of applying all appropriate methods and then minimization in selection was carried out based on the result they output. From the previous researches that had been carried out, identifying suitable classifiers is a very important step. The best classifier has to be picked in order to help in solving a classification problem. Classification in machine learning is also a supervised learning approach wherein the computer program learn from the data being supplied and later make use of the finding to classify new observation (Singh, 2013).

Castro et al. (2017) show a comparative analysis of DNN approaches for crop recognition from multi-temporal sequences of Synthetic Aperture Radar (SAR) images where DNN are compared with a RF classifier. The classifiers are evaluated under different crop dynamics, two-sentinel-1A datasets were used as a testbed. The first, from Germany, is a comparatively simple dynamics of temperate regions and the second dataset is from a tropical region in Brazil. The results from the analysis show that DNN approaches outperformed the RF classifiers.

In choosing the suitable approach for the screening of the disease known as diabetic retinopathy patients in (Mohammadian et al., 2017), a comparison between nine classification algorithms had been performed. These classifiers that were chosen are Neural Network, SVM, Adaptive boosting, K-Nearest neighbours, Gaussian process, Quadratic discriminant analysis, Naive Bayes, RF, and DT. Every classifier are enhanced based on the tuneable parameters and compared together based on accuracy, F1-score, precision and recall. According to the results of the research, the Gaussian process classifier performed the best for distinguishing cases with diabetic retinopathy, meanwhile, SVM and Adaptive boosting approaches can be tuned for a better performance.

### **III. CLASSIFIERS TECHNIQUES**

In this section, four classifications were used in this article and its review provides the formation of hypotheses for the possible relation between data characteristics and method performances.

#### **Deep Neural Network**

Operational models that are supported by DNN is composed of several hidden processing layers that enable it to learn numerous data representations with multi-level abstraction (Prabu, 2018). DNN is a model with a single input and output layer while having an arbitrary number of hidden layers. The values that had been transferred from one layer to the next will actually multiply by weights. The layer was made of neurons that resemble that of human brain's neurons. Figure 1 shows the structure of a DNN. Application of DNN has earned attention in its ability to extract useful features automatically as it is trained to do. The technique introduced in this research has achieved a very reliable predictive accuracy on several occasion where it even rivals human performance (Montavon et al., 2018). To support the claim on DNN, predictive

execution using other classifiers for the sample data are taken into consideration especially in term of accuracy and sensitivity.

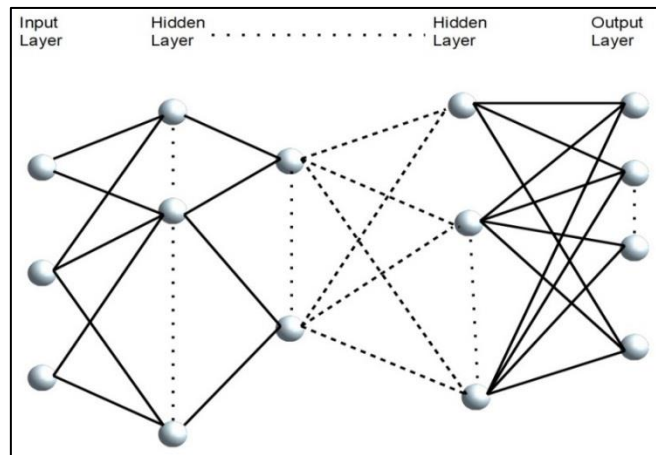


Figure 1. Deep Neural Network (Jeong, 2017)

### Random Forest

The RF developed by (Ali et al., 2012) is a group of un-pruned classification and the regression trees were created by randomly picking training data. RF has a direct connection between the numbers of trees in the forest and according to the output from RF, the bigger the number of trees, the more exact the outcome. In (Breiman, 2001), RF classifier will proceed to bag and utilize random subspace method to create a randomized decision tree that combined the output that produces the best prediction. Figure 2 shows the structure of RF. Classified in the leaf node by a random forest of 3 decision trees is a single instance. Each decision node will be given a unique identifier. After the node passes the test, the corresponding class will then be asserted. Finally, a combination of encoding from each tree is done by making use of concatenation from majority voting then produces final class (Vens & Costa, 2011).

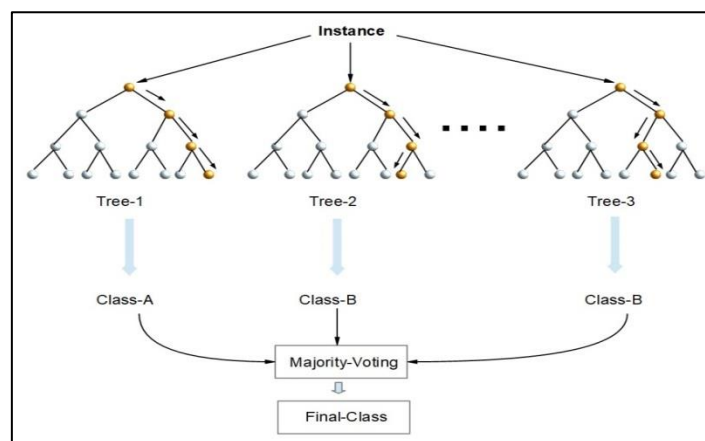


Figure 2. Random Forest (Koehrsen et al., 2018)

### Decision Tree

DT is a guide to a progression of related decisions. It is a model that can be used as a decision-making tool, for research analysis or strategic planning. The processes of finding the root node and splitting the internal nodes make the RF look different than the DT. DT is a supervised classification approach (Zhao & Zhang, 2008) and it is

constructed from internal nodes in the form of circles and the branches of tree segments. A tree segment is what connects between terminal or leaf nodes. It grows downward from the root and drawn from left to right. The grouping data of DT is according to the values of attributes of the given data. A DT structure is as shown in Figure 3.

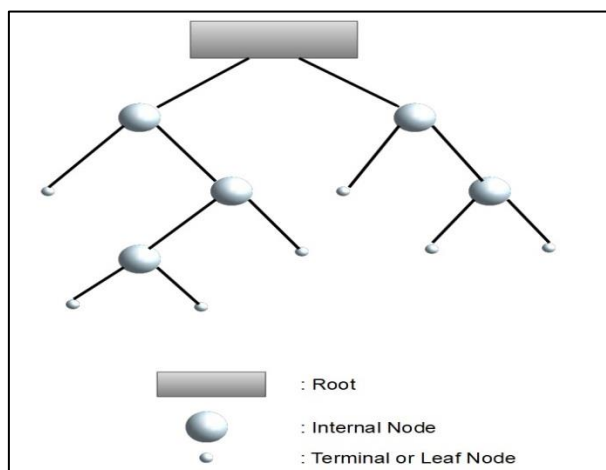


Figure 3. Decision Tree (Ali et al., 2012)

### Support Vector Machine

The purpose of SVM is finding the hyper plane that produces an output of the largest minimum distance to the training data. SVM defines as a supervised machine learning algorithm that is useful for regression and classification. Each data item will be plotted on an  $n$ -dimensional space,  $n$  represents the number of features before performing classification that will provide a distinction between the two classes. A margin is a distance between the hyper plane and the closest data point from either set as shown in Figure 4. Support vectors define the hyper plane while hyper plane separate different classes of the target variable. In order to classify new data more correctly, it is necessary to select a hyper plane that has the greatest margin within any point within the training set.

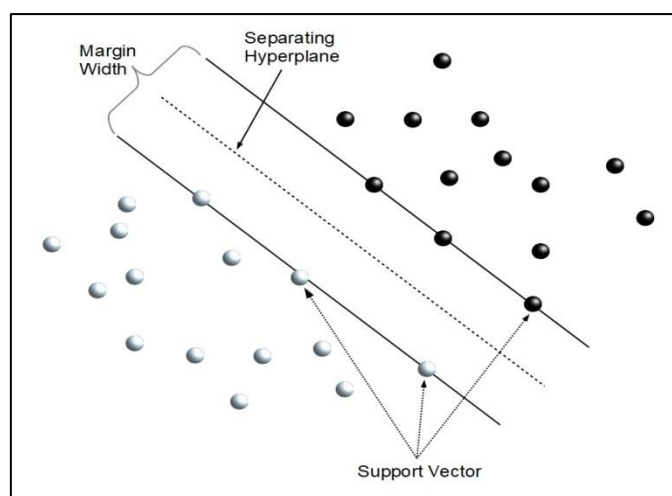


Figure 4. Support Vector Machine (Contacts, 2018)

#### IV. METHODOLOGY

Two processes in classification are known as learning and classification steps (Contacts, 2018; Vens & Costa, 2011; Zhao & Zhang, 2008). Learning step is when the classification algorithm analyzes the training database instances. Meanwhile, the classification step is done by utilizing model which is tested with different datasets that model's classification accuracy.

Through the first step, a classifier is built in order to describe a set of data classes that had already been predetermined. A classification algorithm assembles the classifier by analyzing from a training set that is made up of database tuples and related class label. A class label attribute is a part of database attribute in assuming the predefined class of each tuple. The individual tuples in the training set or training tuples are randomly sampled from the database. In classification, data tuples can be referred as samples, for instances, data points or objects (Han, Kamber, & Pei, 2012). For the second step, the predictive ability of the classifier is evaluated. The precision of a classifier on a given test set is a percentage of the test set that had been accurately classified by the classifier.

The public datasets taken from two Portuguese secondary schools are analyzed in this study. The data set can be referred from UCI machine learning site (Cortez & Silva, 2008). It consists of 650 and 396 number of instances for two subjects: Portuguese and Mathematics with 33 numbers of attributes. While there are 403 instances on student knowledge of Electrical DC Machines subjects with five different of attributes (Kahraman et al., 2013). Therefore, this execution assessment will be performed using two measurements: accuracy and sensitivity.

Performance metrics are utilized in order to measure the generalization ability of the trained classifier (Hossin & Sulaiman, 2015). The performance metrics of four classifiers were determined using accuracy and sensitivity rates. It was common as a metric for accuracy or error rate to be used among researchers in evaluating the generalization ability of classifiers. Accuracy metric is calculated based on the total correctness and the total number of instances are predicted correctly against the unseen data of the trained classifier. Accuracy is the level in effectively classify instances (Oladokun et al., 2008), while sensitivity can be viewed as an ability to differentiate fluctuations in observed event. Sensitivity act as a measurement of the extent of positives that are correctly distinguished.

If the model accuracy is better, the model can be used for classification (Saini & Kumar Jain Asst, 2013). Therefore, the evaluation of the performance of classifier needs to be carried out. Test records are predicted correctly and incorrectly by the classifier will be used for the classifier evaluation and these numbers are arranged in a table known as the confusion matrix (Tan et al., 2006). Confusion matrix supplies the data which is necessary to decide how well a classifier would perform, making it very beneficial for the comparison of the performance of various classifiers. For that matter, a performance metric accuracy is calculated as follows:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

Usually, the classification algorithms will look for classifiers that show the highest accuracy. While sensitivity is calculated as the number of correct prediction with the total number of positives. Below shows the sensitivity rate:

$$\text{Sensitivity} = \frac{\text{Number of correct positive predictions}}{\text{Total number of positives}}$$

## V. EXPERIMENT RESULTS AND ANALYSIS

Four classification algorithms have been used in this work, which popular in the machine learning community (Van Hulse et al., 2007). WEKA or also known as Waikato Environment for Knowledge Analysis, is a data analysis tool which contains the implementation of data pre-preparation, clustering, classification, association rules, visualization and regression of comparison between algorithms (Zhao & Zhang, 2008). The WEKA tool that is being utilized runs RF, DT and SVM model with a cross-validation of 10 folds to avoid overfitting problem (Ali et al., 2012; PrafulKoturwar et al., 2015). The DNN runs 3 hidden layers at 0.3 learning rate and 0.2 momentum. Meanwhile, the RF runs 100 bag size percent, 1 number slot, 1.0 minimum number instances, 0.001 minimum variance for the split, 1 seed and 100 number of iteration. DT used 0.25 confidence factor and 2 minimum number object. While SVM used 3-degree, 0.0 gamma, 40.0 cache size and 0.1 loss.

For training and testing phases of the classifiers, 30 percent of the available data has been selected on randomly as the testing set, whilst the others are being used for the training set. The testing and training sets are saved the equal for all simulations, to maintain the outcomes comparable (Mohammadian et al., 2017).

In order to measure the efficiency of classifiers, accuracy and sensitivity rates are used as the measurement for this study. In general, sensitivity indicates how well the test in predicting a category while accuracy is to gauge how well the test in predicting categories.

The results of the classification reveals that the DNN performs very well in comparison with other classifiers with the highest overall accuracy, followed by RF, SVM and DT. DT performs poorly and less accurate than the others. The result obtained from the experiment of education based case study shows DNN had achieved an accuracy of 99.2% for Mathematics, 96.8% for Portuguese and 96.9% for Electrical DC Machines subject. While, for sensitivity rate, the highest result was 99.2% (DNN) for Portuguese, 96.7% (DNN) for Mathematics and 97.1% (DNN) for Electrical DC Machines subject.

As a conclusion derived from the experiment that had been conducted, different types and number of instances are working best with a certain type of classifier. However, DNN gives better results for accuracy rate for the three different of subjects. Figure 5 and Figure 6 show the percentages of accuracy and sensitivity rates for performance measurement of each classifier.

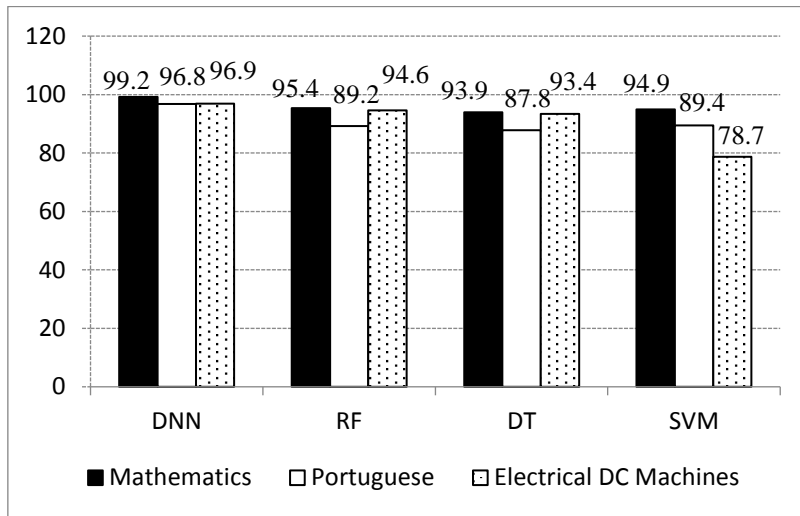


Figure 5. Accuracy rate of the classifiers on the data

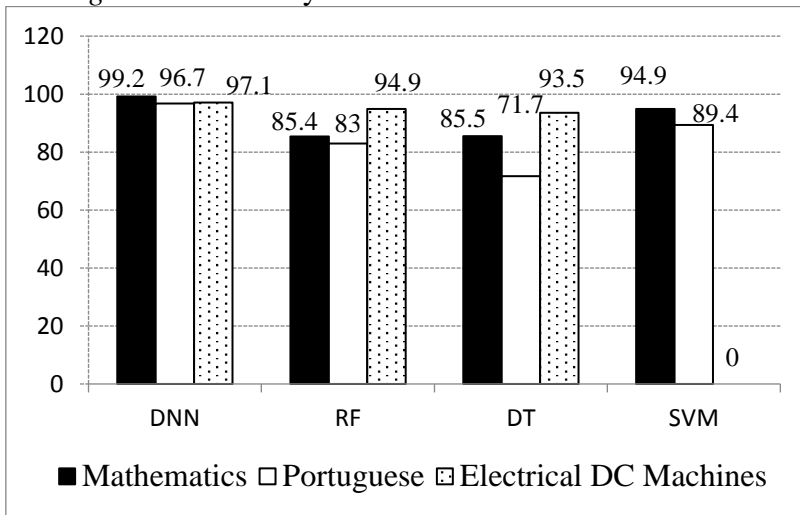


Figure 6. Sensitivity rate of the classifiers on the data

## VI. CONCLUSION

There is a large number of compressive classification algorithms (Duriqi, 2016), but in this paper, a comparative analysis carried out using four classifier algorithms to choose the proper classification algorithm in future research. DNN, RF, SVM, and DT were compared in terms of accuracy and sensitivity rates. The classifiers were tested for three different subjects of education case study, namely Mathematics, Portuguese and Electrical DC Machines. The results show DNN achieve a higher percentage of accuracy and sensitivity rate, thus, it outperform other types of machine learning algorithm. Thus, DNN can be used as a reliable algorithm for predicting in the education sector. Moreover, it can help an understudy's exhibition and recognize understudies who has higher chance of bombing beforehand to give cure. For future work, from these classifiers, the classifiers can hybridize to proceed toward prediction method that able to guide the students to further the research in higher education sector.



## ACKNOWLEDGMENT

This research work is supported by Universiti Teknologi Malaysia under the Research University Grant (GUP), Vote No. 15J11.

## REFERENCES

- Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random forests and decision trees. *IJCSI International Journal of Computer Science Issues*, 9(5), 272–278.
- Bhuyar, V. (2014). Comparative Analysis of Classification Techniques on Soil Data to Predict Fertility Rate for Aurangabad District. *International Journal of Emerging Trends E Technology in Computer Science*, 3(2), 200–203.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Caruana, R., & Niculescu-mizil, A. (2006). An Empirical Comparison of Supervised Learning Algorithms, 161–168. <https://doi.org/10.1145/1143844.1143865>
- Castro, J. B., Feitosa, R. Q., Cue, L., Rosa, L., Diaz, P. A., Del, I., & Sanches, A. (2017). A Comparative Analysis of Deep Learning Techniques for Sub-tropical Crop Types Recognition from Multitemporal Optical / SAR Image Sequences.
- Contacts, A. R. (2018). How-to simulate Support Vector Machine ( SVM ) in R | en . prof . me How-to simulate Support Vector Machine ( SVM ) in R How-to simulate Support Vector Machine ( SVM ) in R | en . prof . me, 4–6.
- Cortez, P., & Silva, A. (2008). Using Data Mining To Predict Secondary School Student Performance. *5th Annual Future Business Technology Conference, 2003(2000)*, 5–12. <https://doi.org/10.13140/RG.2.1.1465.8328>
- Duriqi, R. (2016). Comparative Analysis of Classification Algorithms, 335–338.
- Han, J., Kamber, M., & Pei, J. (2012). 8 - Classification: Basic Concepts. In *Data Mining (Third Edition)* (pp. 327–391). <https://doi.org/http://dx.doi.org/10.1016/B978-0-12-381479-1.00008-3>
- Hossin, M., & Sulaiman, M. N. (2015). Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, 5(2), 1–11. <https://doi.org/10.5121/ijdkp.2015.5201>
- Jeong, D. (2017). Long-Term Prediction of Vehicle Trajectory Based on a Deep Neural Network ¶. *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, 725–727.
- Kahraman, H. T., Sagiroglu, S., & Colak, I. (2013). The development of intuitive knowledge classifier and the modeling of domain dependent data. *Knowledge-Based Systems*, 37, 283–295. <https://doi.org/10.1016/j.knosys.2012.08.009>
- Koehrsen, W., Scientist, D., Student, M., & Communicator, D. S. (2018). Random Forest Simple Explanation Decision Tree : The Building Block, 1–10.
- Mohammadian, S., Karsaz, A., & Roshan, Y. M. (2017). A Comparative Analysis of Classification Algorithms in Diabetic Retinopathy Screening, (Iccke), 631–631. <https://doi.org/10.18293/SEKE2017-207>
- Montavon, G., Samek, W., & Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing: A Review Journal*, 73, 1–15. <https://doi.org/10.1016/j.dsp.2017.10.011>
- Oladokun, V. O., Adebajo, A. T., & Charles-Owaba, O. E. (2008). Predicting students' academic performance using artificial neural network: A case study of an engineering course. *The Pacific Journal of Science and Technology*, 9(1), 72–79.
- Prabu, P. (2018). Students' Performance Prediction Using Deep Neural Network, 13(2), 1171–1176.
- PrafulKoturwar, S., & Mukhopadhyay, D. (2015). A Survey of Classification Techniques in the Area of Big, 1–7.
- Saini, P., & Kumar Jain Asst, A. (2013). Prediction using Classification Technique for the

- Students' Enrollment Process in Higher Educational Institutions. *International Journal of Computer Applications*, 84(14), 975–8887.
- Singh, R. (2013). An Empirical Study of Applications of Data Mining Techniques for Predicting Student Performance in Higher Education. *International Journal of Computer Science and Mobile Computing*, 2(February), 53–57.
- Suthaharan, S. (2014). Big Data Classification : Problems and Challenges in Network Intrusion Prediction with Machine Learning. *Performance Evaluation Review*, 41(4), 70–73. <https://doi.org/10.1145/2627534.2627557>
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Classification : Basic Concepts , Decision Trees , and. *Introduction to Data Mining*, 67(17), 145–205. [https://doi.org/10.1016/0022-4405\(81\)90007-8](https://doi.org/10.1016/0022-4405(81)90007-8)
- Van Hulse, J., Khoshgoftaar, T. M., & Napolitano, A. (2007). Experimental perspectives on learning from imbalanced data. *Proceedings of the 24th International Conference on Machine Learning - ICML '07*, (May 2014), 935–942. <https://doi.org/10.1145/1273496.1273614>
- Vens, C., & Costa, F. (2011). Random forest based feature induction. *2011 IEEE 11th International Conference in Data Mining (ICDM)*, 774–753.
- Zhao, Y., & Zhang, Y. (2008). Comparison of decision tree methods for finding active objects. *Advances in Space Research*, 41(12), 1955–1959. <https://doi.org/10.1016/j.asr.2007.07.020>