DEEP LEARNING APPROACH FOR STUDENT PERFORMANCE PREDICTION IN E-LEARNING

SALMA HUSSEIN OMAR

A thesis submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Computer Science

School of Computing
Faculty of Engineering
Universiti Teknologi Malaysia

AUGUST 2019

ACKNOWLEDGEMENT

First and foremost, I would like to express utmost gratitude to Allah S.W.T for endless blessings and given me strength and knowledge during the development of this research until it has completed.

I am also heartily grateful and thankful to my supervisor Dr. Nor Bahiah HJ Ahmad for her constant support during my study at Universiti Teknologi Malaysia. She inspired me greatly to work in this project. Her willingness to motivate me contributed tremendously to my project. Without her guidance, this thesis would not have been possible and I shall eternally be grateful for the assistance. I have learned a lot from her and I am fortunate to have her as my mentor and supervisor, May Allah bless her and her family.

My sincere appreciation also extends to all my fellow classmates and friends for the willingness to share knowledge and provide assistance when needed. I am also grateful for the love and endless support from my family members especially my mother and father.

Besides, I would like to thank the authority of Universiti Teknologi Malaysia for providing me with a good environment and facilities during my stay in the university.

ABSTRACT

The data in E-learning is generated as a result of the students' interactions during the learning sessions while accessing files and learning resources in E-learning. The learning system collects a huge amount of student's data, such as the registration records, assessment results, and interaction log activities. The enormous amount of data stored in educational databases can provide useful information if analyzed and processed. Predicting the performance of students engaging with the E-learning platform is crucial and has a great impact on both the educational institute and students. This research aims at predicting the performance of students based on Deep Learning (DL) approach using Convolutional Neural Network (CNN). CNN trains a classifier on data by passing learned features through different layers of hidden features. CNN comprises of various layers of which are convolutional layer, pooling layer, activation layer, and fully connected layer. The dataset used in the study is obtained from Open University in the United Kingdom and it has been published in UCI repository. After performing feature selection, 25 attributes have been identified as very significant for student performance prediction using CNN. The training parameters such as the learning rate, weight decay, and optimizer were used to improve the performance of the CNN classifier. In addition to that, the effect that the number of convolutional layers, number of nodes and number of epochs were investigated and compared to evaluate whether it affects the accuracy of the prediction of the classifier. A comparative study being done showed that CNN out-performed Decision Tree and Artificial Neural Network algorithms by giving 72% accuracy, while the other two algorithms obtained 60.46% accuracy and 63.13 %, respectively. In conclusion, the CNN technique proves to be able to predict student academic performance and achieve high accuracy than the other two techniques. For future works, it is suggested to increase the data size, to change the environmental set up from CPU to GPU in order to investigate the performance of deep learning using CNN in large educational data.

ABSTRAK

Kandungan data dalam E-pembelajaran telah dihasilkan daripada interaksi pelajar semasa proses pembelajaran semasa mencapai fail dan sumber pembelajaran dalam sistem pembelajaran. Sistem pembelajaran mengumpulkan data pelajar seperti rekod pendaftaran, keputusan pentaksiran, dan log setiap aktiviti transaksi pelajar. Data yang disimpan dalam pangkalan data pembelajaran menyediakan maklumat yang berguna jika dianalisa dan diproses. Peramalan penilaian prestasi pelajar berdasarkan penglibatan pelajar dalam E-pembelajaran adalah penting dan mempunyai impak yang tinggi terhadap institusi pendidikan dan pelajar. Kajian ini bertujuan untuk meramal prestasi pelajar melalui pendekatan Pembelajaran Mendalam (PM) menggunakan Pengkonvolutan Rangkaian Neural (PRN). PRN melatih pengelas data melalui ciri-ciri yang dipelajari menerusi beberapa lapisan ciri-ciri tersembunyi. PRN mempunyai beberapa lapisan iaitu lapisan pengkonvolutan, lapisan penggemblengan, lapisan pengaktifan, dan lapisan penyambungan penuh. Set data yang digunakan dalam kajian ini didapati daripada Open University di United Kingdom dan telah dikongsi di repositori UCI. Setelah menjalankan pemilihan ciri, 25 attribut telah dikenal pasti mempunyai kesan signifikan terhadap anggaran prestasi pelajar menggunakan PRN. Parameter latihan seperti kadar pembelajaran, pereputan berat, dan pengoptimuman telah digunakan untuk menambah baik prestasi pengelas PRN. Tambahan pula, kesan bilangan lapisan pengkonvolutan, bilangan nod, dan bilangan epok telah dikaji dan dibanding untuk menilai sama ada memberi kesan kepada ketepatan jangkaan pengelas. Kajian komparatif telah dilakukan menunjukkan PRN mengatasi prestasi Pokok Keputusan dan algoritma Jaringan Neural Buatan dengan menunjukkan 72% ketepatan, sementara dua lagi algoritma masing-masing memperoleh 60.46% dan 63.13% ketepatan. Sebagai kesimpulan, teknik PRN telah membuktikan bahawa Tteknik ini mampu untuk meramal prestasi akademik pelajar dengan ketepatan yang tinggi berbanding dua teknik yang lain. Untuk kajian pada masa hadapan, adalah dicadangkan untuk meningkatkan saiz data, mengubah persekitaran daripada CPU kepada GPU bagi mengkaji keberkesanan prestasi pembelajaran mendalam menggunakan PRN dalam data pebelajaran yang besar.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Network
CART	-	Classification and Regression Tress
CNN	-	Convolutional Neural Network
DL	-	Deep Learning
DM	-	Data Mining
DNN	-	Deep Neural Network
DT	-	Decision Tree
DT-GI	-	Decision Tree With Gini Index
DT-IG	-	Decision Tree With Information Gain
EDM	-	Educational Data Mining
E-learning	-	Electronic Learning
ID3	-	Iterative Dichotomiser 3
LA	-	Learning Analytics
LMS	-	Learning Management System
ML	-	Machine Learning
MLP	-	Multilayer Perceptron
NB	-	Naïve Bayes
NN	-	Neural Network
OULAD	-	Open University Learning Analytics Dataset
RNN	-	Recurrent Neural Network
SMOTE	-	Synthetic Minority Over-sampling Technique
UTM	-	Universiti Teknologi Malaysia
VLE	-	Virtual Learning Environment
WEKA	-	Waikato Environment for Knowledge Analysis

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview

Student performance is a great concern for any educational institution to ensure that students perform well in their studies. Predicting the performance of the students and providing appropriate help to the low performing students can improve the quality of the education provided by the institutions. Universities and colleges whose aim is to help improve the quality of the education, the creation of human capital is continually assessed, predicting the students' performance crucial for educational institutions to ensure that the quality of the learning process sufficiently meets the needs of students. Prediction of students' performance focuses on estimating the value of the variable that describes the student, which is unknown. This estimate can be a numerical or categorical value (Jacob *et al.*, 2016).

Universities are currently working in very dynamic and highly viable environments. Students access learning content as learning systems become accessible easily anywhere. In examining the students' online activities such as the mouse clicks, discussion forum, navigating through pages can provide deep insights into how a student performs academically. The tremendous growth of electronic data in various formats like records, files, documents, images, sound, videos, and scientific data in universities maps the need to have some meaningful information extracted from these large volumes of data. The advancement in the data mining field makes it possible to mine educational data to improve the quality of the educational processes. Focusing on and evaluating students' performance at an early stage of their studying period, students with low or high academic performance can be spotted for additional aid (Tegegne and Alemu, 2018). Thus, it became important to gather and analyze students' activities and behaviors to discover learning patterns. The process of understanding students and helping on their academic performance and motivating them to improve

their performance is essential. The mechanisms to explore the gathered data required techniques capable of accurately predicting student achievement as early as possible. Based on the generated predictions, identified students that are at risk to drop out can get the support and assistance they need to improve their learning (Helal *et al.*, 2018).

According to Wolff *et al*. (2014), early prediction of students' behavior can help to make immediate interventions for improving student performance and reduce the dropout or failure rate. In conventional learning, educators have physical interaction and connections with the students to assist them. On the other hand, in online learning students interact with a Virtual Learning Environment (VLE) to access the learning material. Therefore, for educators to know which student needs assistance, educational data mining and learning analytics are applied.

Prediction of students' behavior and discovering prominent factors that can affect students' learning process is of great benefit for any educational institute, aiming to help student performance and determination in completing their studies successfully. Educational data mining enables the education institution to uncover and find hidden pattern based on data collected from E-learning (Stefanova and Kabakchieva, 2017).

E-learning is online learning platforms, which gather records of all student interaction in an online learning setting. With the increase of online learning systems such as E-learning, teaching and learning have become easier than ever before. E-Learning is commonly used in learning environments and it involves learning and teaching through the Internet. With plenty of resources readily available, students can study their preferable subjects through an online class so as to improve their knowledge at their own flexibility. Learning Management System (LMS), E-learning and other similar online learning environments provide flexible learning opportunities to students who are in different locations. These systems generate data stored on log activities, which provides an insight into how students are doing in online learning (Lytras *et al.*, 2018).

The data created in the log files contain crucial information and need to be analyzed to help and understand students' learning process and improve it. Learning analytics, and educational data mining are some of the mechanisms used in mining data and can be applied in all types of education teachings; tradition teaching and online education. Learning analytics, with the data mining techniques interpret the massive data generated to assess progress, predict performance and discover problems students face. Learning analytics provide meaningful insights into the learning pattern of learners and on module assessments.

Analyzing the students' record data together with online activities contributes directly to enhancing students' learning outcome and supporting guidance (Laveti *et al.*, 2017). Rahman and Islam (2018) predicted the academic performance of students using data mining techniques. They used four classification techniques namely; Naïve Bayes, Artificial Neural Network, Decision Tree and K-Nearest Neighbor techniques and also implemented ensemble techniques like Bagging, Adaboosting, and Random Forest to get more accurate results. The used dataset was from E-learning Kalboard 360. Both the studies show that the ensemble technique gives the best accuracy of 84.3 % and Adaboosting on Artificial Neural Network gives 78.6 % accuracy. Most researchers use data mining techniques which are classification and clustering to predict the performance of students. However, lately but with Deep Learning (DL) emerging as an improved technique to study massive datasets in the real world. Few researchers such as (Okubo *et al.*, 2017; Fok *et al.*, 2018) use it for predicting final grades of students. DL is a subset of machine learning techniques and has achieved great success in classifying and pattern recognition of data, such as image analysis, speech recognition and text understanding (Zhang *et al.*, 2018). DL has a number of advantages over traditional data mining techniques, including that the input data includes all data of possible relevance to the prediction problem at hand, nonlinearities and complex interactions among input data are accounted for seamlessly, overfitting is more easily avoided than traditional high dimensional procedure and there exists fast, scale computational frameworks (Tensorflow, Keras, PyTorch, Anaconda).

Okubo *et al.* (2017) used one of Deep Learning technique which is Recurrent Neural Network (RNN) to predict the academic performance of students. Another DL

technique is Convolutional Neural Network (CNN), which comprises of convoluted layers and pooling layers in processing data. Currently, CNN is widely used in classifying images by adjusting the weight and process the data in less complexity.

Due to the advance of E-learning environments, they gather a vast amount of related data to the students in digitized form from various forms in different formats like records, documents, and activity logs. It is crucial to use predictive models in analyzing effectively the massive accumulated data. Therefore, this study focuses on developing a classifier that can predict students' performance based on the students' interactivity with E-learning using DL on CNN.

1.2 Problem Background

E-learning is a popular learning method which uses web or other technologies to promote learning. The process of learning from the internet is called e-learning and the LMS have made it feasible for knowledge to reach almost everywhere (Pant *et al.*, 2018). With the proliferation of the internet and communication technology, E-learning has extensively been used in the process of delivering and acquiring knowledge and has since become a popular learning environment for students. The data generated through online learning is one of the sources for the increase educational data (Moubayed *et al.*, 2018). According to Herath (2018), he stated that education sector is expanding rapidly and which leads to the growth of more educational data to be analyzed.

Education platforms store a huge amount of educational data belonging to the students using it and can be analyzed to improve the learning structures and academic outcomes (Romero and Ventura, 2013). Educational Data Mining is used to extract important information from the repositories containing students' interactions while Learning Analytics is used to understand and to optimize the learning practices, together with the situation in which it occurs (Sciarrone, 2018). Quick growth to the educational data set requires a more sophisticated and advanced set of techniques (Dutt *et al.*, 2017). The ability to identify students at risk of academic failure at an early stage enables a proactive approach to the implementation of strategies improving the quality

teaching and the continuity of students in the teaching and learning process (Alves *et al.*, 2017).

In E-learning, predicting the performance of a student is a great concern to the education managements. For example, it could give an appropriate warning to students who are at risk by forecasting the grade of students, and help them to avoid problems and overcome all difficulties in the study. However, measuring the academic performance of students is challenging because the academic performance of students depends on various factors or characteristics such as demography, personal, educational background, psychological, academic progress and other environmental variables (Guo *et al.*, 2016). Such data stored in the database are student enrolment and attendance records, as well as the examination results. For that reason, mining such data delivers interesting knowledge to the education management.

Analyzing data from E-learning supports and helps students in the learning processes, offering learning materials and educational tools. Moreover, the teachers can follow and manage the resource and monitor the online activity of students. They can use this platform to know and understand more about the students' behavior, based on the data that the platform stores in the databases. In an E-learning environment, students engage with their courses improving the learning outcomes. It increases the ability for students to study on their own and increases their critical thinking skills and self-responsibility (Chitra and Raj, 2018).

Prediction of students' performance is regarded as one of the most important and challenging research topics for educational institutions due to the huge volume of data in the educational repositories (Huang and Fang, 2013; Shahiri *et al.*, 2015). The need for exploiting and analyzing educational data has led to the execution of data mining techniques effectively in studying students' academic behavior and predict their performance (Guo *et al.*, 2016). The early detection of low performers is crucial; this will decrease dropout ratios, improve educational outcomes and provide high-quality education. Educational Data Mining (EDM) mainly benefits when identifying at-risk students, the learning behavior of students, reducing dropouts' rates, and optimizing curriculum (Kostopoulos *et al.*, 2017).

Several types of data EDM and LA have been applied in the process of learning students' success and behavior (Aldowah *et al.*, 2019). They show the capability of classifying a small number of classes with good accuracy. However, data mining main constraint associates to the set of features extracted and used for classification. Although widely known feature extraction may show a good of features, it is a complex task to choose a preliminary set when creating the classifier. Researchers conducted many studies in predicting students' performance using EDM techniques. Classification is the most popular technique used whereby students are classified into certain classes, for example, Excellent, Very Good, Good, Pass (Abu, 2016); Pass or Fail or even grades A, B, C and D (Guo *et al.*, 2016; Asif *et al.*, 2017). Another technique used in predicting students' performance is clustering, students are divided into clusters, and how they perform academically such as High Performers or Low Performers also as Poor or Good (Gowri *et al.*, 2017).

Nowadays deep learning has received a great deal of success in many domains such as image analysis, speech recognition, computer vision, health monitoring, and text understanding (Polson and Sokolov, 2018). It implements both supervised and unsupervised learning, where the features are arranged in hierarchical layers and performs tasks of classification and pattern recognition. DL uses data reduction techniques for creating high dimensional predictors, as it continues to play a vital role in analyzing massive data. Thus, for the past few years, many deep learning models have been created for big data learning. CNN is the most widely used architecture for big data learning and analyzing (Zhang *et al.*, 2018). CNN has the ability to automatically learn all kinds of hidden attributes from the raw data without manually extracting the features.

1.3 Problem Statement

Student performance prediction is expected to help develop certain arrangements that educational institutions organize to improve the performance of students in their studies. As a result of technological development, e-learning platforms provide students with a way to learn the behavior and their interactions with the system through system logs files, data can be analyzed to obtain meaningful

knowledge. With that, education platforms storing a great deal of educational data from the students who use it, the need to analyze the data become obvious to the universities in order to enhance learning structures and academic results. Measuring students ' academic performance is challenging because students ' academic performance depends on different factors or features such as demographics, personal, educational background, psychological, academic and other environmental variables. Through the process of extracting parameters from the database, one has to identify the features which will do the analysis of predicting and improving students' performance (Jacob *et al.*, 2016).

Various traditional EDM techniques have been used to conduct classification and prediction of students' data, such applied techniques are decision tree and association rules. The fast growth of artificial intelligence and deep learning techniques has given another strategy to smart classification and outcome forecast in the latest years. The thesis investigated the performance of CNN technique in predicting the academic performance of students based on the log activities and score grades.

Many researches were done in predicting students' performance and were able to predicate accurately. However, the limitation with educational data mining is mostly it works with small data and algorithms use a set of defined rules such as algorithm are the classification decision tree. In addition to that, the process of selecting features in data mining techniques is aggressively tested to find the best feature to predict students ' performance and accuracy achieved in multi-classification is low compared to binary classification. Due to those issues, a CNN classifier is developed to predict students' performance, as it computes high dimensional data with no need for manual effort.

1.4 Research Questions

The following are the research questions that set the path how this study was conducted:

i. How accurate is to use CNN in predicting students' performance?

ii. How effective is the developed classifier compared to other educational data mining techniques in terms of classification accuracy?

1.5 Research Aim and Objectives

The aim of this thesis is to predict students' academic performance based on their interaction with E-learning using Deep Learning. The objectives of this research are as follows:

i. To develop a classifier that predicts students' performance using CNN in E-learning.

ii. To validate the performance of the developed classifier with educational data mining techniques.

1.6 Scope of the Research

The scopes of this thesis are described below:

i. The boundary of this work focuses on the students' performance prediction based on their interaction with E-learning environment.

ii. The dataset is obtained from Open University containing courses presented at the university. The dataset contains courses, records of students and their assessments results and log activities.

iii. Anaconda, Python is used to develop and implement the classifier.

iv. Weka is used to normalize the datasets.

v. To assess how good the classifier is, the performance metrics; confusion matrix, accuracy score, recall score, precision score from the Python package, Scikit-learn is used to validate.

1.7 Significance of the Research

Predicting the performance of students is important both to the lectures as well as the student. If monitored the performance and learning behaviour of students, it can easily be identified which student needs help in improving their academic performance. Predicting those students that are likely to fail the course can get the best intervention available. In this study, the students' record and log activities are analyzed to get an insight into their learning behaviour and predict the successful outcome. The research is expected to contribute by recommending to use CNN technique to develop a classifier that can predict the performance of students in E-learning compared to EDM techniques. Furthermore, it will assist in knowing that the features studied in this thesis can help to understand the behavior of students and used to assess their final grades.

1.8 Thesis Organization

This thesis is organized into five chapters. Chapter 1 provides information regarding problems and significant of the study. Chapter 2 contains a review of student performance prediction and existing techniques while explaining the previous concepts and opinions from experts or authors. This chapter presents with previous work or existing models for research. Chapter 3 presents the thesis methodology by introducing the architecture of the developed classifier and how the objectives are achieved. Chapter 4 discusses the data preprocessing and the development of the classifier. Finally, in the last chapter, summary, and conclusions of the presented study is given as well as some suggestions for future research.

REFERENCES

Abdul Razak, R., Omar, M., Ahmad, M. and Mara, P. (2018) 'A Student Performance Prediction Model Using Data Mining Technique', 7, pp. 61–63.

Abu, A. (2016) 'Educational Data Mining & Students' Performance Prediction', *International Journal of Advanced Computer Science and Applications*, 7(5), pp. 212–220.

Abubakar, Y. and Ahmad, N. B. H. (2017) 'Prediction of Students ' Performance in E-Learning Environment Using Random Forest', *International Journal of Innovative Computing*, 7(2), pp. 1–5.

Adejo, O. W. and Connolly, T. (2018) 'Predicting student academic performance using multi-model heterogeneous ensemble approach', *Journal of Applied Research in Higher Education*, 10(1), pp. 61–75.

Al-Saleem, M., Al-kathiry, N., Al-osimi, S. and Badr, G. (2015) 'Machine Learning and Data Mining in Pattern Recognition 11th International Conference, MLDM 2015 Hamburg, Germany, July 20-21, 2015 Proceedings', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9166, pp. 403–414.

Al-Shehri, H., Al-Qarni, A., Al-Saati, L., Batoaq, A., Badukhen, H., Alrashed, S., Alhiyafi, J. and Olatunji, S. O. (2017) 'Student performance prediction using Support Vector Machine and K-Nearest Neighbor', *Canadian Conference on Electrical and Computer Engineering*, pp. 17–20.

Aldowah, H., Al-Samarraie, H. and Fauzy, W. M. (2019) 'Educational data mining and learning analytics for 21st century higher education: A review and synthesis', *Telematics and Informatics*. Elsevier, 37(April 2018), pp. 13–49.

Alves, P., Miranda, L. and Morais, C. (2017) 'The Influence of Virtual Learning Environments in Students' Performance', *Universal Journal of Educational Research*, 5(3), pp. 517–527.

Amutha, P. and R. Priya, D. (2018) 'A survey on educational data mining techniques in predicting student's academic performance', *International Journal of Engineering & Technology*, 7(3.3), p. 634.

Annika Wolff, Zdenek Zdrahal, D. H. and Knoth, P. (2014) 'Predicting Student Performance

from Combined Data Sources', *Studies in Computational Intelligence*, 524.

Asif, R., Merceron, A., Ali, S. A. and Haider, N. G. (2017) 'Analyzing undergraduate students' performance using educational data mining', *Computers & Education*. Elsevier Ltd.

Avella, J. T., Kebritchi, M., Nunn, S. G. and Kanai, T. (2016) 'Learning analytics methods, benefits, and challenges in higher education: A systematic literature review', *Journal of Asynchronous Learning Network*, 20(2).

Bara, M. W., Ahmad, N. B., Modu, M. M. and Ali, H. A. (2018) 'Self-organizing map clustering method for the analysis of e-learning activities', *Proceedings of Majan International Conference: Promoting Entrepreneurship and Technological Skills: National Needs, Global Trends, MIC 2018*, pp. 1–5.

Bendangnuksung and Prabu, D. (2018) 'Students' Performance Prediction Using Deep Neural Network', *International Journal of Applied Engineering Research*, 13(2), pp. 1171–1176.

Bharara, S., Sabitha, S. and Bansal, A. (2018) 'Application of learning analytics using clustering data Mining for Students' disposition analysis', *Education and Information Technologies*. Education and Information Technologies, 23(2), pp. 957–984.

Cantabella, M., Martínez-España, R., Ayuso, B., Yáñez, J. A. and Muñoz, A. (2019) 'Analysis of student behavior in learning management systems through a Big Data framework', *Future Generation Computer Systems*. Elsevier B.V., 90, pp. 262–272.

Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2002) 'SMOTE: Synthetic minority over-sampling technique', *Journal of Artificial Intelligence Research*, 16(January), pp. 321–357.

Chitra, A. P. and Raj, M. A. (2018) 'E-Learning: Challenges and Research Opportunities Using Machine Learning & Data Analytics', 3, pp. 11–13.

Costa, E. B., Fonseca, B., Almeida, M., Ferreira, F., Araújo, D. and Rego, J. (2017) 'Computers in Human Behavior Evaluating the effectiveness of educational data mining techniques for early prediction of students ' academic failure in introductory programming courses', *Computers in Human Behavior*. Elsevier Ltd, 73, pp. 247–256.

Daud, A., Aljohani, N. R., Abbasi, R. A., Lytras, M. D., Abbas, F. and Alowibdi, J. S. (2018) 'Predicting Student Performance using Advanced Learning Analytics', pp. 415–421.

Devasia, T., Vinushree, T. P. and Hegde, V. (2016) 'Prediction of students performance using Educational Data Mining', *Proceedings of 2016 International Conference on Data Mining and Advanced Computing, SAPIENCE 2016*, pp. 91–95.

Dubey, A., Kamath, S. and Kanakia, D. (2016) 'Learning Data Mining Techniques', *International Journal of Computer Applications*, 136(11), pp. 5–8.

Dutt, A., Ismail, M. A. and Herawan, T. (2017) 'A Systematic Review on Educational Data Mining', *IEEE Access*, 5(January), pp. 15991–16005.

Fok, W. W. T., He, Y. S., Yeung, H. H. A., Law, K. Y., Cheung, K., Ai, Y. and Ho, P. (2018) 'Prediction model for students' future development by deep learning and tensorflow artificial intelligence engine', *2018 4th International Conference on Information Management, ICIM 2018*, pp. 103–106.

Gowri, G. S., Thulasiram, R. and Baburao, M. A. (2017) 'Educational Data Mining Application for Estimating Students Performance in Weka Environment', *IOP Conference Series: Materials Science and Engineering*, 263(3).

Guo, B., Zhang, R., Xu, G., Shi, C. and Yang, L. (2016) 'Predicting Students Performance in Educational Data Mining', *Proceedings - 2015 International Symposium on Educational Technology, ISET 2015*. IEEE, pp. 125–128.

Hamsa, H., Indiradevi, S. and Kizhakkethottam, J. J. (2016) 'Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm', *Procedia Technology*, 25, pp. 326–332.

Hasibur Rahman, M. and Rabiul Islam, M. (2018) 'Predict Student's Academic Performance and Evaluate the Impact of Different Attributes on the Performance Using Data Mining Techniques', *2nd International Conference on Electrical and Electronic Engineering, ICEEE 2017*. IEEE, (December), pp. 1–4.

Hegde, V. (2018) 'Higher Education Student Dropout Prediction and Analysis through Educational Data Mining', *2018 2nd International Conference on Inventive Systems and Control (ICISC)*. IEEE, (Icisc), pp. 694–699.

Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D. J. and Long, Q. (2018) 'Predicting academic performance by considering student heterogeneity', *Knowledge-Based Systems*. Elsevier, 161(December 2017), pp. 134–146.

Herath, D. (2018) 'Educational Data Mining to Investigate Learning Behaviors : A Literature Review', (May).

Hidaka, A. and Kurita, T. (2017) 'Consecutive Dimensionality Reduction by Canonical Correlation Analysis for Visualization of Convolutional Neural Networks', *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications*, 2017(0), pp. 160–167.

Hong, B., Wei, Z. and Yang, Y. (2017) 'Discovering Learning Behavior Patterns to Predict

Dropout in MOOC', (Iccse), pp. 700–704.

Hu, X., Cheong, C. W. L., Ding, W. and Woo, M. (2017) 'A Review on Predicting Student's Performance Using Data Mining Techniques', *Procedia Computer Science*, 72, pp. 414–422.

Huang, S. and Fang, N. (2013) 'Computers & Education Predicting student academic performance in an engineering dynamics course : A comparison of four types of predictive mathematical models', *Computers & Education*. Elsevier Ltd, 61, pp. 133–145.

Hussain, S., Dahan, N. A., Ba-Alwib, F. M. and Ribata, N. (2018) 'Educational data mining and analysis of students' academic performance using WEKA', *Indonesian Journal of Electrical Engineering and Computer Science*, 9(2), pp. 447–459.

Jacob, J., Jha, K., Kotak, P. and Puthran, S. (2016) 'Educational Data Mining techniques and their applications', *Proceedings of the 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015*, pp. 1344–1348.

Jan, B., Farman, H., Khan, M., Imran, M., Islam, I. U., Ahmad, A., Ali, S. and Jeon, G. (2019) 'Deep learning in big data Analytics: A comparative study', *Computers and Electrical Engineering*. Elsevier Ltd, 75, pp. 275–287.

Kavitha, G. and Raj, L. (2017) 'Educational Data Mining and Learning Analytics Educational Assistance for Teaching and Learning', *International Journal of Computer & organization Trends*, 41(1), pp. 21–25.

Kesavaraj, G. and Sukumaran, S. (2013) 'A Study On Classification Techniques in Data Mining', *IEEE*.

Khanna, L., Singh, S. N. and Alam, M. (2016) 'Educational Data Mining and its Role in Determining Factors Affecting Students Academic Performance : A Systematic Review'.

Khanna, L., Singh, S. N. and Alam, M. (2017) 'Educational data mining and its role in determining factors affecting students academic performance: A systematic review', *India International Conference on Information Processing, IICIP 2016 - Proceedings*.

Kim, K. G. (2016) 'Deep Learning Book Review', *Healthcare Informatics Research*, 22(4), pp. 351–354.

Kostopoulos, G., Lipitakis, A., Kotsiantis, S. and Gravvanis, G. (2017) 'Predicting Student Performance in Distance Higher Education Using Active Learning', *18th International Conference on Engineering Applications of Neural Networks, EANN 2017*, 1(Dm), pp. 75–86.

Kularbphettong, K. (2018) ‘Applied Computational Intelligence and Mathematical Methods’, 662.

Kuzilek, Jakub; Hlosta, Martin; Herrmannova, Drahomira; Zdrahal, Zdenek and Wolff, A. (2015) ‘OU Analyse : Analysing at - risk students at The Open University’, *Learning Analytics Review, LAK15-1*.

Laveti, R. N., Kuppili, S., Ch, J., Pal, S. N. and Babu, N. S. C. (2017) ‘Implementation of learning analytics framework for MOOCs using state-of-The-Art in-memory computing’, *Proceedings - 2017 5th National Conference on E-Learning and E-Learning Technologies, ELELTECH 2017*, (1).

Li, L. Y. and Tsai, C. C. (2017) ‘Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance’, *Computers and Education*. Elsevier Ltd, 114(300), pp. 286–297.

Liang, J., Yang, J., Wu, Y., Li, C. and Zheng, L. (2016) ‘Big Data Application in Education : Dropout Prediction in Edx MOOCs’, *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)*. IEEE, pp. 440–443.

Liu, T., Fang, S., Zhao, Y., Wang, P. and Zhang, J. (2015) ‘Implementation of Training Convolutional Neural Networks’.

Lytras, M. D., Aljohani, N. R., Visvizi, A., Ordonez De Pablos, P. and Gasevic, D. (2018) ‘Advanced decision-making in higher education: learning analytics research and key performance indicators’, *Behaviour and Information Technology*. Taylor & Francis, 37(10–11), pp. 937–940.

Maulida, J. D. and Kariyam (2017) ‘Students academic performance based on behavior’, *AIP Conference Proceedings*, 1911.

Mhetre, V. (2017) ‘Classification based data mining algorithms to predict slow, average and fast learners in educational system using Weka.’, (Iccmc), pp. 475–479.

Moubayed, A., Injadat, M., Nassif, A. B., Lutfiyya, H. and Shami, A. (2018) ‘E-Learning: Challenges and Research Opportunities Using Machine Learning Data Analytics’, *IEEE Access*. IEEE, 6, pp. 39117–39138.

Namratha, B. (2016) ‘Educational data mining –applications and techniques’, *International Journal of Latest Trends in Engineering and Technology*, 7(2), pp. 484–488.

Nguyen, N. G., Tran, V. A., Ngo, D. L., Phan, D., Lumbanraja, F. R., Faisal, M. R., Abapihi, B., Kubo, M. and Satou, K. (2016) ‘DNA Sequence Classification by Convolutional Neural Network’, *Journal of Biomedical Science and Engineering*, 09(05), pp. 280–286.

Nurhayati, O. D., Bachri, O. S., Supriyanto, A. and Hasbullah, M. (2018) 'Graduation prediction system using artificial neural network', *International Journal of Mechanical Engineering and Technology*, 9(7), pp. 1051–1057.

Okubo, F., Yamashita, T., Shimada, A. and Ogata, H. (2017) 'A neural network approach for students' performance prediction', (February 2018), pp. 598–599.

Pant, V., Bhasin, S. and Jain, S. (2018) 'Self-Learning system for personalized E-Learning', *2017 International Conference on Emerging Trends in Computing and Communication Technologies, ICETCCT 2017*, 2018-Janua, pp. 1–6.

Pereira, B. A., Pai, A. and Fernandes, C. (2017) 'A Comparative Analysis of Decision Tree Algorithms for Predicting Student's Performance', *International Journal of Engineering Science and Computing*, 7(4).

Polson, N. G. and Sokolov, V. O. (2018) 'Deep learning', *Nature Methods*, 13(1), p. 35.

Pradeep, A., Das, S. and Kizhakkethottam, J. J. (2015) 'Students Dropout Factor Prediction Using EDM Techniques', *2015 International Conference on Soft-Computing and Networks Security (ICSNS)*. IEEE, pp. 1–7.

Qu, H. and Chen, Q. (2015) 'Visual analytics for MOOC data', *IEEE Computer Graphics and Applications*. IEEE, 35(6), pp. 69–75.

Rani, U. K., Ramadevi, N. G. and D, L. (2016) '2016 International Conference on Computing for Sustainable Global Development (INDIACom)', pp. 1623–1627.

Romero, C. and Ventura, S. (2013) 'Data mining in education', 3(February), pp. 12–27.

Roy, S. and Garg, A. (2018) 'Predicting academic performance of student using classification techniques', *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics, UPCON 2017*, 2018-Janua, pp. 568–572.

Ruby, J. and David, K. (2014) 'Predicting the Performance of Students in Higher Education Using Data Mining Classification Algorithms - A Case Study', *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 2(Xi), pp. 173–180.

Sciarrone, F. (2018) 'Machine learning and learning analytics: Integrating data with learning', *2018 17th International Conference on Information Technology Based Higher Education and Training, ITHET 2018*. IEEE, pp. 1–5.

Shahiri, A. M., Husain, W. and Rashid, N. A. (2015) 'A Review on Predicting Student's Performance Using Data Mining Techniques', *Procedia Computer Science*. Elsevier Masson SAS, 72, pp. 414–422.

Shruthi, P. and Chaitra, B. P. (2016) ‘Student Performance Prediction in Education Sector Using Data Mining’, *International Journal of Advanced Research in Computer Science and Software Engineering*, 6(3), pp. 212–218.

Sivakumar, S., Venkataraman, S. and Selvaraj, R. (2016) ‘Predictive modeling of student dropout indicators in educational data mining using improved decision tree’, *Indian Journal of Science and Technology*, 9(4), pp. 1–5.

Song, D. (2018) ‘Learning Analytics as an Educational Research Approach Learning Analytics as an Educational Research Approach’, (October).

Sreenivasa Rao, K., Swapna, N. and Praveen Kumar, P. (2017) ‘Educational data mining for student placement prediction using machine learning algorithms’, *International Journal of Engineering & Technology*, 7(1.2), p. 43.

Stefanova, K. and Kabakchieva, D. (2017) ‘Educational data mining perspectives within university big data environment’, *2017 International Conference on Engineering, Technology and Innovation: Engineering, Technology and Innovation Management Beyond 2020: New Challenges, New Approaches, ICE/ITMC 2017 - Proceedings*, 2018-Janua, pp. 264–270.

Tegegne, A. K. and Alemu, T. A. (2018) ‘Educational data mining for students’ academic performance analysis in selected Ethiopian universities’, *Information Impact: Journal of Information and Knowledge Management*. African Journals Online (AJOL), 9(2), p. 1.

Umadevi, S. and Marseline, K. S. J. (2018) ‘A survey on data mining classification algorithms’, *Proceedings of IEEE International Conference on Signal Processing and Communication, ICSPC 2017*, 2018-Janua(July), pp. 264–268.

Venkatachalapathy, K., Vijayalakshmi, V. and Ohmprakash, V. (2017) ‘Educational data mining tools: A survey from 2001 to 2016’, *Proceedings - 2017 2nd International Conference on Recent Trends and Challenges in Computational Models, ICRTCCM 2017*, (14), pp. 67–72.

Vieira, C., Parsons, P. and Byrd, V. (2018) ‘Visual learning analytics of educational data: A systematic literature review and research agenda’, *Computers and Education*. Elsevier Ltd, 122, pp. 119–135.

Wong, B. T. M. (2017) ‘Learning analytics in higher education: an analysis of case studies’, *Asian Association of Open Universities Journal*, 12(1), pp. 21–40.

Zacharis, N. Z. (2016) ‘Predicting Student Academic Performance in Blended Learning Using Artificial Neural Networks’, *International Journal of Artificial Intelligence &*

Applications, 7(5), pp. 17–29.

Zacharis, N. Z. (2018) 'Classification and Regression Trees (CART) for Predictive Modeling in Blended Learning', *International Journal of Intelligent Systems and Applications*, 10(3), pp. 1–9.

Zaffar, M., Hashmani, M. A. and Savita, K. S. (2018) 'Performance analysis of feature selection algorithm for educational data mining', *2017 IEEE Conference on Big Data and Analytics, ICBDA 2017*, 2018-Janua, pp. 7–12.

Zainudin, Z., Shamsuddin, S. M. and Hasan, S. (2019) 'Deep Learning for Image Processing in WEKA Environment', 11(1), pp. 1–21.

Zaki, M. J. and Meira, W. J. (2013) *Data Mining and Analysis: Fundamental Concepts and Algorithms*, *Cambridge University Press*.

Zhang, Q., Yang, L. T., Chen, Z. and Li, P. (2018) 'A survey on deep learning for big data', *Information Fusion*. Elsevier, 42(October 2017), pp. 146–157.