Predicting Students At-Risk Using Deep Learning Neural Network

Tshepang Motaung
School of Computer Science
and Applied Mathematics
The University of the Witwatersrand
Johannesburg, South Africa
1431795@students.wits.ac.za

Dr. Ritesh Ajoodha
School of Computer Science
and Applied Mathematics
The University of the Witwatersrand
Johannesburg, South Africa
ritesh.ajoodha@wits.ac.za

***Abstract*—The availability of educational data, along with technology-enhanced learning platforms, allows for the extraction of students' learning behavior, mitigation of their worries, optimization of the educational environment, and data-driven decision-making. To detect at-risk pupils, this study uses a deep learning neural network trained on a set of attributes taken from a learner activity tracker application called experience API (xAPI). Other probabilistic models such as Random Forest, Support Vector Machine, and k Nearest Neighbours were compared to the convolutional neural network and multilayer perceptron models in correctly classifying at-risk students, and information gain ranking was used to eliminate features that are said to be redundant or irrelevant. After 10 fold cross-validation, the Support Vector Machine model outperformed all other models, including the Convolutional Neural Network and the Multilayer Perceptron, with a peak accuracy score of 77.51%, while the convolutional neural network and the multilayer perceptron achieved accuracy scores of $72.58\%\ and\ 73.73\%$, respectively. Student behavioral features, according to the research, are useful predictors of student performance.**

***Index Terms*—Deep learning, Behavioral factors, Classification, Student Performance Prediction, Educational Data Mining (EDM)**

I. Introduction

Rapid advances in technology-enhanced learning platforms have led to massive growth in online educational data and an adequate education repository has been generated with a substantial impact on higher education institutions. The buildup of educational data has sparked the formation of numerous academic communities, including learning analytics, which anticipates learner behavior and provides indicators for optimal policy formulations [1].

Academic performance in students is difficult to assess since it is influenced by a range of factors including demographics, educational experience, and other external influences. Machine learning approaches have recently become popular for investigating educational data and identifying hidden relevant patterns for forecasting students' grades. Deep learning, which emerged from machine learning is defined by multiple computational layers, and allows the model to learn from instances, patterns, or events ([3], [4]), displacing traditional ways of hand-engineering the features. Despite the fact that numerous models have been studied in the context of learning analytics research, determining the utility of deep learning in learning analytics is still in its development, and research into its implementation has only just emerged [2]. Learning analytics entails collecting, aggregating, reviewing, and analyzing student data in order to gain a better knowledge of the learning environment and improve the performance of students and instructors [1]. This research compares the best probabilistic machine learning models from other researchers to the implemented deep learning neural network and analyses the students' accomplishments in the context of learning students' behavioral attributes.

The following is how the rest of the paper is structured: Section II of the paper discusses relevant efforts in EDM. The data collection and preprocessing are discussed in Section III. The proposed methodology is presented in Section IV. Section V details the experiments and their outcomes. Finally, in Section VI, we sum up this article with some closing observations and suggestions for future work.

II. Related Work

The broad recognition that improving student academic achievement can have significant economic and social benefits, means improving students academic performance is important for economic and social development [10].

According to author [6], socioeconomic status (SES) has an impact on students academic progress, and this has long been established in sociological studies. While there has been an increase in research on predicting student academic performance, there are few empirical studies on socioeconomic determinants and other factors that may influence educational outcomes among students from low socioeconomic backgrounds. Author [6] acknowledges that there is endless information about the link between family socioeconomic status and the student academic success, and the factors that could impact educational results within particular socioeconomic status bands have not been thoroughly assessed. [6] established factors like sex, age, ethnicity, unexplained absences, housing type and parental educational attainment to be crucial factors and predictors towards students academic performance. Language is another impediment to students' performance in post secondary institutions, according to author [7] of literature on background characteristics.

Author [8] developed a dropout abstract model that emphasized the importance of casual faculty interactions. The model interacted with institutional image, size, admissions, and other predicted background variables. These institutional characteristics were expected to have an impact on informal communication with teachers, educational outcomes (e.g. academic performance and career aspirations) and other varsity experiences (e.g. peer culture and extracurricular). Informal interaction with teachers and other varsity experiences were predicted to influence educational outcomes. Author [9] investigated students' academic achievement on an e-learning (10-week introduction level) approach and used neural networks to predict students' success. The model was used to forecast the final score of students and divide them into two implicit classes corresponding to their performances. The multiple choice test outcomes were the networks set inputs.

Deep learning (DL) assumes that complicated methods may be constructed by reorganizing simple methods such as Gaussian kernels [11]. DL is made up of numerous two layer sequences, a feature detection layer, and a feature pooling layer, each with a supervisory step [11]. Author [12] used supervised and unsupervised deep learning to build prediction models with supplied inputs and outputs, as well as to extract useable attributes from raw data, such as internet services that employ learning algorithms to extract relevant data from data acquired on the internet.

The Table I below present the accuracy's and models obtained by other author's:

TABLE I
ACCURACIES OBTAINED BY OTHER RESEARCHERS.

Authors	Model	Accuracy
[14]	ANN	73.8%
[1]	SVM	85.65%
[7]	RF	73%
[8]	MLP	88.2%
[15]	Decision Tree	65%

In conclusion, different studies have been conducted to use data mining approaches to solve educational difficulties. Few research, on the other hand, have looked into student conduct during the process of learning and its effect on academic performance.
This paper will look at the effect of pupils engagement with the e-learning system. Furthermore, the information gathered will assist schools in improving student performance. Additionally, to assist administrators in improving learning systems.

Research Hypothesis 1: Deep learning models outperform probabilistic models in predicting at-risk students.

Research Hypothesis 2: Demographic features such as family status, age, gender, etc have an influence in students academic performance.

III. DATA COLLECTION AND PREPARATION

The data for this research was gathered utilizing the Experience API from the Kalboard 360 E-Learning system (XAPI) [13] and [14]. Kalboard is a multi-agent learning management system (LMS) that uses cutting-edge technologies to make learning easier [13]. With such a system, users can access instructional content in real time from any device with an Internet connection [14]. The goal of X-API in this study is to trace student habits throughout the educational process in order to assess elements that may influence a student's academic progress.
The collection contains 480 student records with 16 attributes. The traits are grouped into 3 categories: (1) Demographic factors include nationality and gender. (2) Academic history characteristics like section, grade stage and educational stage. (3) Personality traits like raised hand in class, resource opening, parents replying survey, and school satisfaction. The attributes of the dataset are listed in Table II, along with their descriptions. As stated in the table, a new feature category, behavioral factor, has been added. During the educational process, this type of feature is linked to learning background and learners conduct.

TABLE II
ATTRIBUTES AND THEIR DESCRIPTIONS

Attributes	Attributes Description	Attributes Categories
Visited Resources	numeric(0-100)	
Absent Days	nominal(above-7/below)	
Raise Hands	numeric(0-100)	
Announcement checked	numeric(0-100)	Behavioural
Parental participation	nominal(Yes/No)	
Discussion	numeric(0-100)	
Parents satisfaction	nominal(Yes/No)	
Country of Birth	nominal(USA,Iraq,etc)	
Relation	nominal(mom/father)	Demographic
Birthplace	nominal(USA,Iraq,etc)	
Gender	nominal(Male/Female)	
Topic	nominal(IT,Math,etc)	
Grade Id	nominal(G-01–12)	
Semester	nominal(1st/2nd)	Academic
Stage Id	nominal	
Section	nominal(A,B,C)	

Our dataset is a multivariate dataset with no missing values. The sample included 305 males and 175 females. 179 pupils represent Kuwait, 172 pupils represent Jordan, 28 pupils represent Palestine, 22 pupils represent Iraq, 17 pupils represent Lebanon, 12 pupils represent Tunis, 11 pupils represent Saudi Arabia, 9 pupils represent Egypt, 7 pupils represent Syria, 6 pupils represent the United States, Iran, and Libya, 4 pupils represent Morocco, and one pupil represents Venezuela. The dataset was gathered during two academic semesters, with 245 pupils marks collected in the first semester and 235 pupils marks obtained in the second. A school attendance component is also included in the data set, which splits students into two groups based on their absence days: There are 191 pupils who have had more than seven absence days, while 289 have had fewer than seven. This dataset also includes a brand-new educational characteristic called parent parturition. The parent engagement feature includes two subfeatures: Parent answering survey and parent school satisfaction. Two hundred and seventy (270) parents replied to the survey, while 210 did not. 292 parents are satisfied with their children's education, while 188 are not.

The data preprocessing stage follows, which entails remodelling the obtained data into a relevant format. We observe from Table II that our dataset consists of both numerical values and categorical values. Since machine learning models are designed to work with numerical inputs, we used One-Hot-Encoding to transform categorical values into numerical inputs. One-Hot-Encoding was chosen over Label-Encoding because our categorical values have no natural ordering. As a preprocessing step our inputs features were scaled to be in the range $[0, 1]$.

IV. Feature Selection

Feature selection enables us to eliminate irrelevant and redundant data, which can cut down on time complexity, enhance learning accuracy, and make the learning model and data easier to understand. By assessing the dependency of a target variable Y on a predictor variable X, and keeping features that have a strong correlation with Y, we used information gain ranking to eliminates features. Table III shows the results of information gain ranking, in which characteristics were retrieved and eliminated from our dataset and then employed with our five off-the-shelf classifiers.

TABLE III
TABLE OF FEATURES MAINTAINED AND FEATURES ELIMINATED

Features Maintained	Eliminated Features
ParentAnsweringSurvey_{No}	Nationality_{SaudiArabia}
ParentAnsweringSurvey_{Yes}	Nationality_{Syria}
Parent_{Mom}	Nationality_{Tunis}
Parent_{Father}	Nationality_{USA}
AbsenceDays>7	Nationality_{Morocco}
AbsenceDays<7	Nationality_{lebanon}
ParentschoolSatisfaction_{Bad}	Nationality—_{Iran}
ParentschoolSatisfaction_{Good}	Nationality_{venzuela}
Gender_{F}	Nationality_{Lybia}
Gender_{M}	Nationality_{Egypt}
VisITedResources	Nationality_{Jordan}
RaisedHands	Nationality_{Iraq}
AnnouncementsView	
Discussion	

Figure 1 depicts the link between the amount of features we supplied our benchmark model and the accuracy we obtained as we added more features to it. After introducing 10 features, we found that our benchmark model's cut-off point was reached when accuracy reached a pick value of 71%.

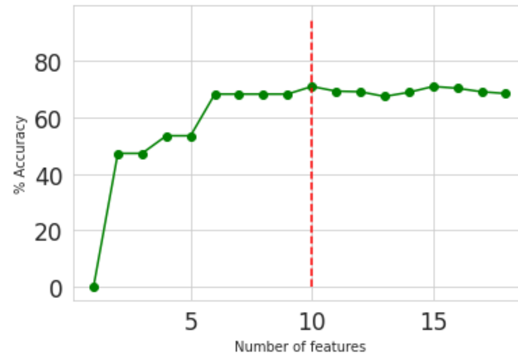

Fig. 1. Classification Accuracy

Our models were able to forecast outcomes based solely on significant attributes as a result of this. As a result, instead of the 16 features in the original dataset, each new entry can now only have 10 features.

The five off-the-shelf classifiers we used are kNearest Neighbor (kNN), support vector machine (SVM), multilayer perceptron (MLP), convolutional neural network (CNN) and Random Forest (RF).

kNN: is a conventional non-parametric classifier [16]. The kNN classifier analyses the distances between the point and points in the training data set to identify an unknown event represented by some feature vectors as a point in the feature space [16]. The on-line'searching' for the k nearest neighbours of a particular testing case is the key computation in kNN, it does not require the off-line training stage. [16]. We used the euclidean distance function to calculate the distance between two points, which has the following format:

$$dist(A, B) = \sqrt{\frac{\sum_{i=1}^{m}(x_i - y_i)^2}{m}}$$

where, $A = (x_1, x_2, ..., x_m)$ and $B = (y_1, y_2, ..., y_m)$ are feature vectors and m is the dimensionality of the feature space [16].

SVM: are supervised learning models that use linked learning algorithms to analyze data for regression analysis and classification. SVM is built on the concept of finding the optimal hyperplane for separating variables into distinct domains. The radial basis function kernel is used in the implemented SVM in the following structure:

$$K(X_1, X_2) = exp(-\frac{||X_1 - X_2||^2}{2\sigma^2})$$

where, σ is our variance and hyperparameter, and $||X_1 - X_2||$ is the Euclidean distance bounded by two points X_1 and X_2.

MLP: is an acyclic graph with a finite number of nodes [17]. The nodes are neurons that have been activated logistically. A node that is not the destination of a connection is called an input neuron, a node that is not the source of a connection is called an output neuron, and all nodes that are neither input nor output neurons are called hidden neurons [17]. The corresponding element of the input pattern is displayed for each input neuron as $a_i \leftarrow x_i$. After calculating the values a_j for all predecessors $j \in Pred(i)$, calculate net_i and a_i for all hidden and output neurons i as follows [17]:

$$net_i \leftarrow w_{i0} + \sum_{j \in Pred(i)} (w_{ij} a_j)$$

where w_{i0} is the bias weight of neuron i, w_{ji} is the weight of the connection $i \rightarrow j$ and $Pred(i)$ is the set of all neurons j for which a connection $j \rightarrow i$ exists.

$$a_i \leftarrow f_{log}(net_i)$$

where f_{log} is the logistic function.
Neurons with $tanh$ activation function are calculated as follows:

$$a_i = tanh(net_i) = \frac{e_i^{net} - e^{-net_i}}{e_i^{net} + e^{-net_i}}$$

CNN: is one kind of feedforward neural network. CNN's structure consists of two layers in general [18]. One is the feature extraction layer, which connects each neuron's input to the previous layer's local receptive fields and extracts the local feature [18]. The positional relationship between the local features and other features will be determined once the local features have been extracted [18]. The other is the feature map layer, which is made up of a number of feature maps for each computing layer of the network [18]. Author [18] defined the model's forward pass format as shown below:
In the output of row i and jth neuron in the lth hide layer H:

$$O_{(l,i)} = \tanh(\sum_{j=0}^{H} \mathbf{O}_{(l-1,j)} \mathbf{W}^l_{(i,j)} + \mathbf{Bias}^{(l,i)})$$

RF: [19] defined as a Decision Tree-Based Classifier that uses voting to select the best classification tree as the algorithm's final classification. It uses the bootstrap approach to generate several subsets of samples, then creates a Decision Tree for each subset of samples and merges several Decision Trees into a Random Forest [19]. When the categorization sample is complete, a vote on the Decision Tree determines the classification's ultimate outcome [19]. Figure 2 represent the training process of the RF of [19].

V. Student Performance Classification

Here we give experimental results from performing student academic performance classification on the xAPI. The classification techniques used in this paper to identify at-risk students were k-Nearest Neighbor(kNN), Random Forest(RF), Support Vector Machine(SVM), Multilayer Perceptron(MLP) and Convolution Neural Network(CNN). The experimental results of the classification utilizing the pruned feature set in Table III are shown in Table IV. Table V lists the various classifier parameters that were employed in the studies.

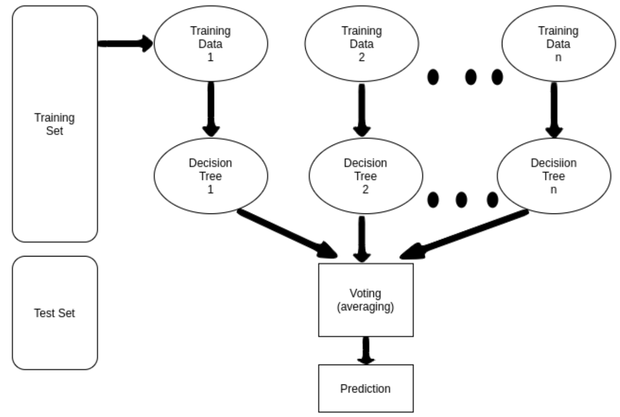

Fig. 2. Flowchart for Random Forest Training [19]

TABLE IV
Students Academic Performance With Pruned Features

Classifier	Accuracy	Time to build the model
kNN	75.74%	0.075 sec
RF	77.36%	0.084 sec
SVC	77.51%	0.010 sec
MLP	73.73%	0.088 sec
CNN	72.58%	2.00 sec

TABLE V
Student Academic Performance Classification Classifier Parameters

Classifier	Parameters
kNN	k=15, with euclidean distance metric, weighting=distance
RF	split_function=entropy, no of trees=50, max_depth=10
SVM	radial basis function kernel, tolerance=1e-6, regularization=1
MLP	hidden layer=3, learning rate=constant, activation=tanh,
CNN	activation=tanh, epoch=20, batch size=10, input shame=(size of the input, 1)

The Support Vector Machine Model produces the best classification score of 77.51% as shown in Table III. The Random Forest and k Nearest Neighbor models performed similarly to the Support Vector Machine and trail behind it, respectively. The Multilayer Perceptron is the classifier that comes after the three previously mentioned classifiers, with the Convolution Neural Network being the classifier that performs the least. Below are the confusion matrix for the two benchmark approaches as well as two of our deep learning models.

The row and column labels denote class labels, with L indicating low-level, M indicating Middle-level, and H indicating High-level.

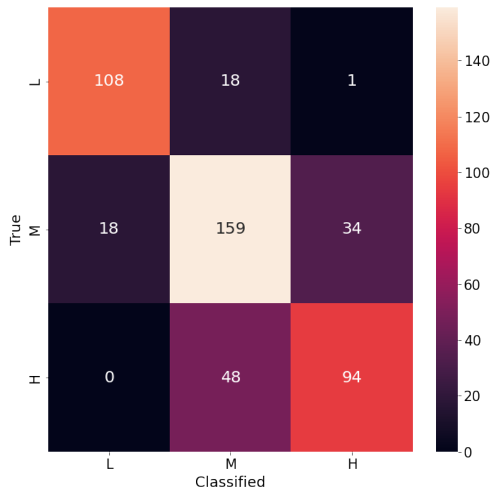

Fig. 3. Using a SVM model with k=10 cross validation

Figure 3 confusion matrix present that we have 108 correctly classified instances in label L, 159 correctly classified instances in label M, and 94 correctly classified instances in label H. Therefore summing up to 361 correctly classified instances using SVM model after 10 fold cross validation.

The RF algorithm uses the bootstrap approach to generate several subsets of samples, then creates a decision tree for each subset of samples and merges several decision trees. When the categorization sample complete, a vote on the Decision tree determines the classification's ultimate outcome [19]. Figure 4 present that we have 110 correctly classified instances in label L, 154 correctly classified instance in label M, and 96 correctly classified instances in label H, according to the confusion matrix. As a result, after 10 fold cross validation, the RF model accurately classified 360 cases.

MLP is a splitting algorithm that can learn models in real time as well as non-linear models. Multi-label classification is supported by the method. The logistic function processes the raw output for each class. Values greater than or equal to 0.5 are rounded to 1, otherwise, they are rounded to 0. For a sample's anticipated output, the indices with the value 1 denote the sample's assigned class. Figure 5 show that we have 102 correctly classified cases in L, 153 correctly classified cases in M, and 79 correctly classified instances in H. As a result, after 10 fold cross validation, MLP model accurately classified 334 instances.

Multiple layers of artificial neurons make up CNN. Artificial neurons are functions that calculate the activation value from a weighted sum of various inputs and outputs. Typically, the first layer extracts basic features. The output is forwarded to the next layer, which looks for more complicated features. It

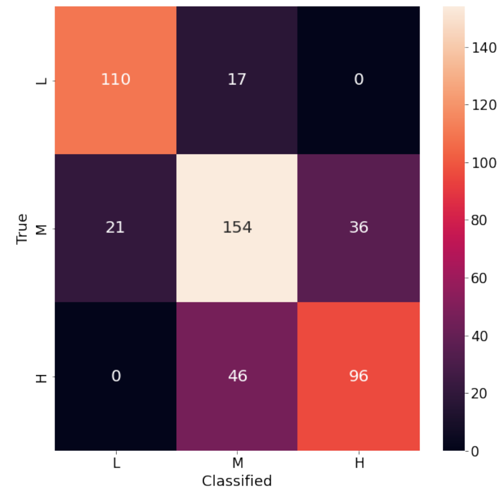

Fig. 4. Using a RF with k=10 cross validation

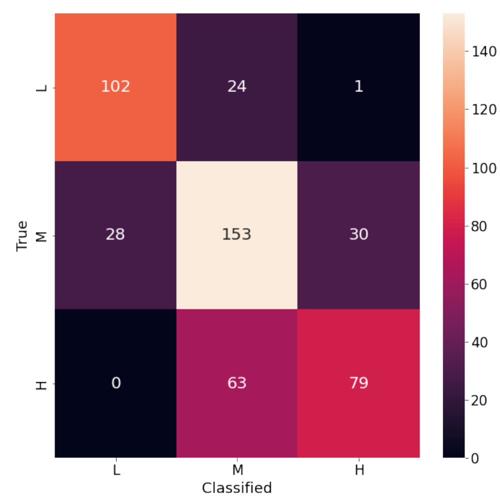

Fig. 5. Using a MLP with k=10 cross validation

can detect even more complicated traits as we proceed further into the network. The classification layer generates a set of confidence scores based on the activation map of the final convolutional layer, which indicate how probable the instance is to belong to the class. Figure 6 present that we have 113 correctly classified cases in L, 123 correctly classified cases in M, and 107 correctly classified cases n H. As a result, after 10 fold cross validation, CNN model accurately classified 343 instances.

VI. Conclusion and Recommendations

In this work, the classification of student academic performance was accomplished by first identifying and extracting factors that influence student performance. Students' features were analyzed for their significance in the task of performance categorization, in order to make the feature vector concise, comprehensive and efficient. After experimenting with a variety of off-the-shelf classifiers, Support Vector Machines and Random Forest proved to be the most accurate models, with

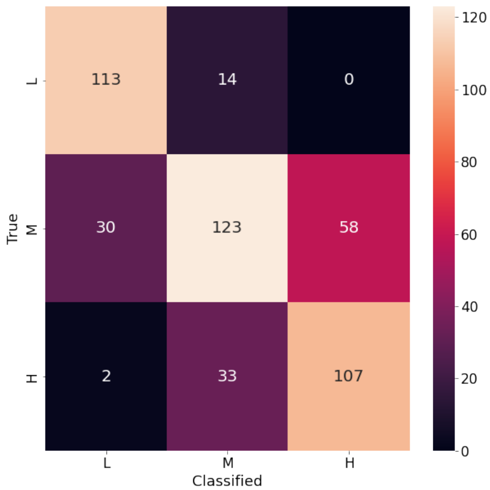

Fig. 6. Using a CNN model with k=10 cross validation

accuracy of 77.51% and 77.36% respectively, outperforming the deep learning models Multilayer Perceptron and Convolutional Neural Network which obtained the accuracy of 73.73% and 72.58% respectively. The experimental results in this study are comparable to benchmarks established by a number of authors. This is one of the first studies to link student behavior to academic achievement. Using data preprocessing techniques, we discovered that behavioral aspects have a positive impact on students' academic success. The learned knowledge obtained through the use of categorization techniques indicates that the learner's actions had a significant role in the learning process, as seen by the high accuracy findings.

The lack of data and a trustworthy ground truth are two obvious drawbacks that may be limiting model effectiveness in categorizing student performance. To generate more accurate answers, we'll use data mining techniques and deep learning models on a larger data set with more distinguishing properties in the future.

References

[1] Waheed, Hajra and Hassan, Saeed-Ul and Aljohani, Naif Radi and Hardman, Julie and Alelyani, Salem and Nawaz, Raheel, "Predicting academic performance of students from VLE big data using deep learning models," Elsevier, vol. 104, pp. 106–189, 2020

[2] Coelho, Orlando Bisacchi and Silveira, Ismar, "Deep learning applied to learning analytics and educational data mining: A systematic literature review," Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE), vol. 28, pp.143, 2017

[3] Nawaz, Raheel and Thompson, Paul and Ananiadou, Sophia "Identification of Manner in Bio-Events.,"Lrec, pp.3505–3510, 2012

[4] Ananiadou, Sophia and Thompson, Paul and Nawaz, Raheel, "Enhancing search: Events and their discourse context,"International conference on intelligent text processing and computational linguistics, pp.318–334, 2013

[5] Khalil, Mohammad and Ebner, Martin, "Learning analytics: principles and constraints,"EdMedia+ Innovate Learning, pp.1789–1799, 2015

[6] Considine, Gillian and Zappalà, Gianni, "he influence of social and economic disadvantage in the academic performance of school students in Australia,"Journal of sociology, vol.28, pp.129–148, 2002 [Sage Publications Sage CA: Thousand Oaks, CA]

[7] Mathye, Macdaline Raisibe, "A theoretical model to predict undergraduate attrition based on background and enrollment characteristics,"2020

[8] Huang, Shaobo and Fang, Ning, "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models,"Computers & Education, vol.61, pp.133–145, 2013 [Elsevier]

[9] Lykourentzou, Ioanna and Giannoukos, Ioannis and Mpardis, George and Nikolopoulos, Vassilis and Loumos, Vassili, "Early and dynamic student achievement prediction in e-learning courses using neural networks,"Journal of the American Society for Information Science and Technology, vol.60 , pp.372–380 , 2009 [Wiley Online Library]

[10] Ajoodha, Ritesh and Jadhav, Ashwini, "Identifying at-risk undergraduate students using biographical and enrollment observations for mathematical science degrees at a south african university,"International Journal of Science and Mathematics Education, pp.1–21 , 2017

[11] Bengio, Yoshua and LeCun, Yann and others, "Scaling learning algorithms towards AI,"Large-scale kernel machines, vol.34, pp.1–41, 2007

[12] Fan, Cheng and Xiao, Fu and Zhao, Yang, "A short-term building cooling load prediction method using deep learning algorithms,"Applied energy, vol.195, pp.222–233, 2017 [Elsevier]

[13] Amrieh, Hamtini and Aljarah, "Mining Educational Data to Predict Student's academic Performance using Ensemble Methods," International Journal of Database Theory and Application, pp.119-136, 2016

[14] Amrieh, Hamtini and Aljarah,"Preprocessing and analyzing educational data set using X-API for improving student's performance,"In Applied Electrical Engineering and Computing Technologies (AEECT), pp.1-5, 2015 [EEE Jordan Conference]

[15] Akour, Mohammed and Alsghaier, Hibal and Al Qasem, Osama, "The effectiveness of using deep learning algorithms in predicting students achievements,"Indonesian Journal of Electrical Engineering and Computer Science,vol.19,pp.387–393,2020

[16] Hu, Li-Yu and Huang, Min-Wei and Ke, Shih-Wen and Tsai, Chih-Fong, "The distance function effect on k-nearest neighbor classification for medical datasets,"SpringerPlus, vol.5, pp.1–9, 2016[SpringerOpen]

[17] Riedmiller, Martin and Lernen, AM,"Multi layer perceptron,"Machine Learning Lab Special Lecture, University of Freiburg, pp.7–24, 2014

[18] Tianyi Liu, Shuangsang Fang, Yuehui Zhao, Peng Wang, Jun Zhang,"Implementation of Training Convolutional Neural Networ,"CoRR, vol.abs/1506.01195, 2016[dblp computer science bibliography, https://dblp.org]

[19] Mohsin Abdulazeez, Adnan and Zeebaree, Diyar and Zebari, Dilovan and Jijo, Bahzad,"Machine Learning Classifiers Based Classification For IRIS Recognition,"Qubahan Academic Journal, vol.1, 2021[10.48161/qaj.v1n2a48]