

Multidimensional Approach Based on Deep Learning to Improve the Prediction Performance of DNN Models

<https://doi.org/10.3991/ijet.v14i02.8873>

Mohammed El Fouki^(✉), Noura Aknin, K. Ed El. Kadiri
Abdelmalek Essaadi University, Tetouan, Morocco
melfouki@uae.ac.ma

Abstract—The most of collected data samples from E-learning systems consist of correlated information caused by overlapping input instances, which decrease the classifier credibility and reliability. This paper presents an improved classification model based on Deep Learning and Principal Component Analysis (PCA) method as its use in reducing the dimensionality of data. By this task, we introduce the best learning process to extract just the useful parameters that describe students' performances in an E-learning system. One of the primary goals of this technique is to help earlier in detecting the dropouts and discovering of students who need special attention, so that the teachers could provide the appropriate counseling at the right time. This study presents the proposal approach and its algorithms. In addition, it shows how deep neural network was modeled in the training phase, and how PCA helps in the elimination of correlated information in our dataset to increase the classifier performance. Finally, we introduce an example of an application of the method in a data mining scenario, find out more references for further information.

Keywords—Educational Data Mining (EDM), Classification, Deep Neural Network (DNN), Deep Learning, Principal Component Analysis (PCA).

1 Introduction

Big Data Analytic, especially, Deep Learning is an interesting area of data science. Big Data has become more important because many public and private organizations have collected massive amounts of information that can contain useful knowledge on different topics such as medical informatics, marketing, fraud detection, cyber security, prediction and learning systems [1]. Consequently, companies such as Microsoft and Google are analyzing large volumes of data, affecting existing and future technologies. The deep learning algorithm is one of these new technologies that try to extract high-level of complex abstractions and representations of data through a hierarchical learning process. These complex abstractions are learned according to several simple abstractions developed in the previous level of the hierarchy. Therefore, this study aims to analyze and learn the massive amounts of unattended data by using deep learning, which is based on an effective algorithm for data analytics where raw data is largely unmarked and unclassified [2]. Deep Learning is a promising algorithm of machine

learning research in extracting complex data characteristics and representations at high levels of abstraction [3]. These algorithms develop a hierarchical schema of learning and data representation, where higher level (more abstract) characteristics are defined as set of lower level characteristics (less abstract).

The quality of the classifier model has an important impact on the performance of the machine learning: a poor model is likely to reduce the performance even for a dataset with good representation, whereas an advanced learning model can lead to a high performance for a relatively complex learner problem [4]. Also, dimensionality reduction techniques, which focuses on building data representations and useful features from input data, is an important technique for optimizing machine learning models. In this paper we present an optimized approach for classification models based on dimensionality reduction technique and deep learning to prove the outperformance of these techniques for modeling and predicting the student's performance in an e-learning system. We started by representing the input data in the training process, then, we developed a model which represents the best domain of interest and predicts the output classes within the higher accuracy. We will also discuss some aspects of Deep Learning model optimization that require further exploration process to achieve specific challenges introduced by machine learning and data analytics, including continuous transmission data, scalability of models.

2 Case Study

Currently, there is a great interest in performances and real success of e-learning educational systems. Hence, many learning programs are now enrolled in a distance learning course offered via the Internet. Therefore, several questions about the true quality of such environments still remain and need to be answered in this regard [5]. Some educational institutions usually academic, have made the leap from learning technologies, in particular the use of higher training tools. On the other hand, just a few empirical studies have examined the personal factors that predict the students' preferences for these learning systems [6].

The main objective of our study is to evaluate the predictive performance of a parametric forecasting methods in a comparative perspective models to predict the learners' performance. Thus, our basic assumption is that, overall, deep parametric methods perform better than a non-parametric prediction. This research stands out by using new techniques known as dimensionality reduction and deep learning to prove the outperformance of these techniques for modeling and predicting the student's performance in an e-learning system. To test our hypothesis, we have resorted to a database of an e-learning platform operated by Abdelmalek Essaadi University for a distance learning public. This e-learning system aims to offer a high-level training in Management Informatics, thus, a platform of a virtual learning environment is set up. It hosts all the contents of the training, and offers to the various project actors (professors, students and administrators) a set of tools encouraging fruitful exchanges. Therefore, on the basis of the collected information we compared the results of discriminant multivariable analysis models.

3 Related Work

Deep Learning models have achieved remarkable results in various machine learning applications, including speech recognition [7], computer vision [8], pattern recognition [9], data analytics [2], robotics [10], self-driving cars [11], Voice generation [12], Natural Language Processing [13], recommendation systems [14], bioinformatics [15], mobile advertising and visual art processing [16].

Classification systems have been used a lot to predict students' performances, for example [17] introduces a genetic algorithm (GA) approach to classify educational web system students based on some information extracted from data recorded in the learning process. Authors improved the prediction accuracy, simply, by adjusting the appropriate weighting characteristics used via the genetic algorithm. Thus, the implemented GA approach has been demonstrated to successfully enhance the accuracy of combined classifier performance. In the same context [18] proposed an automatic method to register computed tomography (CT) and magnetic resonance (MR) brain images by using first principal direction of feature images. In this paper, the PCA method and the neural network algorithm are used to select the first principal directions from feature images. Another paper [19] described an optimized artificial neural network model for hourly prediction of building electricity consumption. They used two different historical datasets for this investigation, collected from the Energy Prediction Shootout Contest.

On the other side, Cireşan, et al [20] described a new approach that achieved a better-than-human recognition rate of 99.46% and won the final phase of the German traffic sign recognition benchmark. Their method was developed based on deep neural network algorithms in a fast, fully parameterizable GPU implementation. Another paper conducted by [21] presented a deep neural network classifier for malware detection based on more than 400,000 software binaries drawn from different customers and some internal malware databases. The model achieves a 95% detection rate at 0.1% false positive rate.

PCA is one of the most widely implemented algorithms in machine learning. It aims to reduce the dimensionality and interpretable linear combinations of the data, but retain as much as possible the original variability in the data. This technique was first introduced by Karl Pearson in 1901, but he did not introduce the practical calculation method for two or more variables, which were worthless for many applications [22]. But since 1930, PCA was involved in several machine learning problems including:

- Limiting the number of variables to be measured.
- Modeling methods such as linear regression, logistic regression or discriminant analysis.
- Identifying homogeneous groups of observations, or, on the contrary, atypical observations.

4 Multidimensional Approach

4.1 Data preparations

The dataset used in this study was obtained from a real word problem. Thus, the input instances were selected from a distance-learning platform operated by Abdelmalk Essaadi University. The initial size of the data is 496 records. In the first step, all the data collected from different tables were joined in a single table and saved as an ARFF file. Then, the final grade of students in the dataset is classified into excellent, very good, good, acceptable and fail based on :

- Students Sections
- First Test Marks
- Selection Marks
- Second Test Marks
- Number of Connections
- Interview Score
- Students' participation in the Forum

4.2 Data selection and transformation

At this stage, just the relevant fields were selected to build our prediction model, thus, only a few derived variables have been chosen. Then we categorized the selected attributes into five classes to simplify the modulization task. While a few learning attributes were extracted from the dataset, all the predictor, input, and output variables which were derived are listed in Table 1.

Table 1. Student Related Variables

Variable	Name	Values
Section	Students Sections	{Licence1, Licence2, Master1, Master2, master mathematical}
Test 1	First Test Marks	{Good, Average, Poor}
SLT	Selection Marks	{Excellent, Acceptable, Very Good, Good, Poor}
Test 2	Second Test Marks	{Good, Average, Poor}
Ncox	Number of Connections	{Good, Acceptable}
Interview	Interview Score	{Good, Average, Poor}
Forum	Students' participation in the Forum	{Good, Average, Poor}
FG	Final grade	{Excellent, Acceptable, Very Good, Good, Fail}

4.3 Data pre-processing

When a large number of quantitative variables are studied simultaneously, it is difficult to represent a global graph for all features. The difficulty arises from the fact that the studied individuals are no longer represented in a space of two dimensions, but in a larger dimension space. E-learning data outcome of this case study is often unbalanced,

inconsistent, and missing in particular behaviors or even trends, which is an emergent problem for many data mining techniques. Therefore, a pre-processing step was needed here as a means to convert the raw measurement attributes into more reduced and more optimized values that the models can more easily handle, and to speed up the convergence process. For all this reason, a dimensionality reduction technique based on principal component analysis was implemented, to compact the dataset and preserve the greatest amount of information from our learning features.

PCA represents the variance-covariance matrix that we used to produce this relevant summary by analyzing the dispersion of the considered data. From this matrix, we have extracted the factors that we need by using a simple mathematical process, which make it possible to produce the desired graphs within small dimensions. It is the interpretation of these features that will enhance our learning process and to better understand the structure of the data analyzed. This process would be guided by a number of numerical and graphic indicators, which are there to help the user make the most accurate and objective interpretation.

4.4 DNN: Deep neural network

Multi-layer neural networks define a class of functions which are able to approach any continuous function with compact support. No other type of neural network will be studied in this work, and thereafter any neural network will be considered as a multi-layer. This definition is inspired from the human biological neuron, the weights are playing the role of synapses, the vector x is the inputs and W is the coefficients or the weights [23]. The function f is called the transfer function or the threshold function.

The p -input neuron is a function $f: \mathbb{R}^{p+1} \times \mathbb{R}^p \rightarrow \mathbb{R}$ within:

$$g: \mathbb{R} \rightarrow \mathbb{R}$$

$$W \in \mathbb{R}^{p+1}, w = (\omega_1, \dots, \omega_{p+1})$$

$$\forall x \in \mathbb{R}^p, f(w, x) = g\left(\sum_{i=1}^p \omega_i x_i + \omega_{p+1}\right)$$

The number of neural network layers is not limited. Networks of two layers (input layer and output layer) are rarely used. Therefore, four layers were needed (input layers, two hidden layers, and the output layer) to build models with an interesting density and property.

4.5 Network training and validation process

The next step in modeling the neural network model is the resolution of the number of processing items and hidden layers in the network. Choosing the number of neurons, activation function and hidden layers are very important, because having a large number of hidden layers/neurons will increase the training time and slow down the learning

process. Similarly, a small number of hidden layers/neurons in a neural network decrease the processing capabilities of the network [24].

Consequently, defining the number of hidden layers is a delicate decision. However, there are two ways to choose the adequate network sizes: growing method which starts with a small network and then increase the model dimension. Pruning method, which begins with several layers in the network and then decrease the model dimension. During the modulization of the prediction model we used the growing method.. Therefore, the model was initialized with no hidden layer then we adjusted the number of layers until we fixed it at two hidden layers of four neurons each (Figure 1).

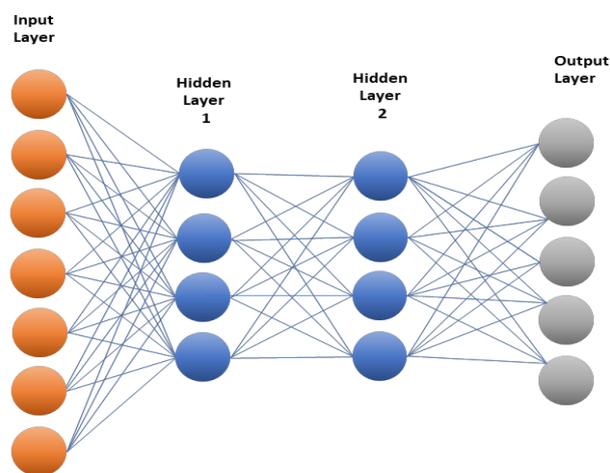


Fig. 1. The Deep Neural Network model

5 Experiments

5.1 Experimental set-up

The algorithm presented in this study begins by preprocessing the e-learning dataset, hence, it was preceded by a preprocessing step using PCA, and followed by transforming the highly correlated input variables to provide a small number of orthogonal variables and improve training speeds. The data were then divided into three datasets; training, validation and tests. The training set was used to train the DNN, while the validation process and the test set were used to evaluate the DNN model performance after achievement of the training process.

5.2 Deep neural network optimization

Nodes in the input layer are being connected to the nodes of the hidden layer through weighted connections. Also, each of the input layer and the two hidden layers are trained with a weight adjustment algorithm as an auto-encoder, in order to reproduce

the input. Basically, by using this structure, in each layer we tried to reproduce the input, and by making this happen with fewer neurons than the input layer, this forces the hidden layer neurons to become good feature detectors.

To develop a custom prediction model with higher accuracy we parameterized our deep learning model based on many input parameters:

- Training sets: select the frame used to build the model. We select randomly 43% from our instances.
- Validation sets: we specified a frame of 29% of our instances to evaluate the accuracy of the model.
- Test sets: we specified a frame of 28% of our instances for predictions.
- nfolds: we used a nfolds equal to 0 to avoid cross-validation.
- Weights_column: we select the most important variable for bias correction.
- Offset_column: Specify a column to use as the offset.
- Checkpoint: we have used this option to optimize the module by used a deep learning process.
- Activation: Specify the activation function (we used Tanh).
- Hidden: Two hidden layers.
- Epochs: the number of times or stream to iterate the dataset.
- Autoencoder: Deep Learning autoencoder.
- Loss function: Specify Absolute as loss function.

5.3 Learning for students' performances classification

In order to have meaningful results it was imperative to test our hypothesis on a real-world problem, thus, our methodology was examined to predict the student's performance based on certain input variables selected from a e-learning platform database operated by Abdelmalek Essaadi University for a distance learning public. For the ease of evaluation, the final grade has been normalized to excellent, acceptable, very good, good and fail. Several measures were applied in this study to investigate our models. Firstly, the Margin curve which defines the points illustrating the prediction margin. The margin curve described in Figure 2, 3 and 4 is a variation between the probability expected for the actual class and the highest probability expected for the other categories. Figure 2 illustrates that the most of the instances are correctly classified by deep neural network model, since they are organized in the area of probability one (the right part of the graph).

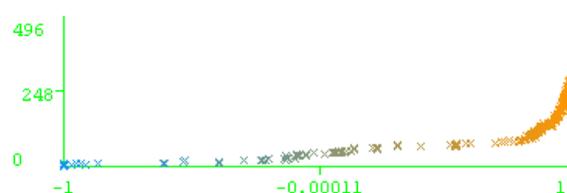


Fig. 2. Margin Curve of DNN Model

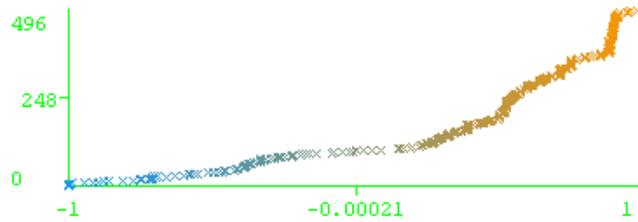


Fig. 3. Margin Curve of MLP Model

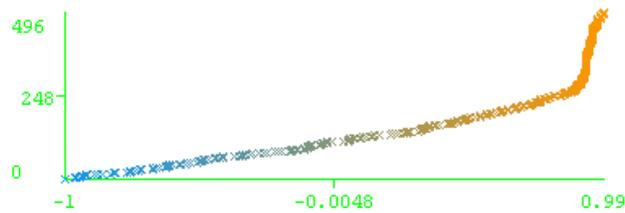


Fig. 4. Margin Curve of Bayes Net Model

Basically, the visualized curves (in figures 2, 3 and 4) show the outcome of adjusting the probability threshold above which a subject is related to that class. Each accuracy presented in WEKA explorers represents one point on the ROC curve. This proves how instances are perfectly classified, which represents the classifier accuracy. So, more the curve (margin curve) is towards the right part of the graph more the classification capabilities increase. These curves are initialized by plotting TPR (the rate of true positives, percentage of instances correctly classified) against FPR (the rate of false positives, the percentage of instances or items incorrectly classified).

The performance measures precision, TP Rate, recall, and F-measure in figure 5 confirm how our deep neural network model, Bays Net classifier, and the MLP classifier perform in identifying output for each instance. The performance measures “accuracy” determines the number of well-classified instances, as an absolute value, or a percentage of the total examples number.

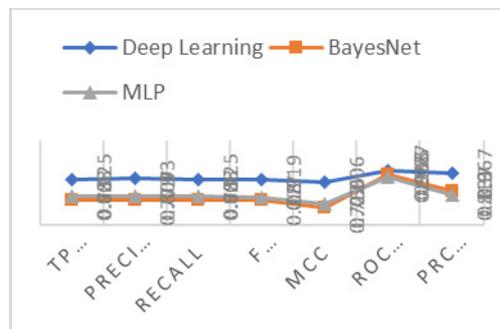


Fig. 5. Measures comparison between BaysNet, Deep Learning and MLP classifiers.

Moreover, the mean absolute error described in table 2 is the difference between the probability (calculated by the classifier) of an instance class, and its initial probability class, which has been fixed in the dataset. The sum of these errors is then divided by the number of instances. It is clear that our model had the best accuracy, by comparing all five measures for the three classifiers we distinct advantage of the optimized deep learning model.

Table 2. Performance comparison between BaysNet, Deep Learning and MLP classifier

	Bays Net	Deep Learning	MLP
Accuracy	78.0242 %	92.5403 %	80.8468 %
Kappa statistic	0.7162	0.9031	0.7506
Mean absolute error	0.1054	0.0412	0.1206
Root mean squared error	0.2429	0.1426	0.2422
Relative absolute error	34.0475 %	13.3088 %	38.9717 %
Root relative squared error	61.7492 %	36.2401 %	61.5811 %

6 Conclusion

The finals grade of students in the dataset are classifieds into excellent, very good, good, acceptable and fail with an accuracy of 92.5%, which is better than the record accuracy of predicting student performance in the literature. The hierarchical architecture of our Deep Learning model is optimized by imitating the deep-layer learning process, which automatically extracts features and abstractions from the underlying data. We started by developing a mining model, and then apply a filter to the input data in order to eliminate the correlated information. Then, we compared the results of these different models using an elevation curves graph. Finally, we extracted additional knowledge from the underlying mining structure. This leads to optimize learning features and to reduce the classification time and even improve the performance of deep neural networks, particularly in resolving complex problems involving a large number of input data.

Experimental studies in this paper have shown that the data representations drown from the stacking of long non-linear extractors as in deep neural networks often give better results than bayesian network and multilayer perceptron. Moreover, The overall prediction accuracy from our analysis (92.5%) is better than the record accuracy of analyzing undergraduate students' performance 83.65% [25], predicting students' performance (between 52-67 %) from Kabakchieva's study [26] and 72.38% from Ramesh's analysis [27]. Therefore, the fact to add a new phase of preprocessing to reduce the dimension of data decreased the correlated information caused by overlapping input instances, which reduced the network training time and improved the performance of DNN systems. However, more work is necessary on how we can implement deep learning algorithms on issues related to scalability of deep learning models, criteria for extracting real data representations, and domain adaptation. Our perspectives include Implementing filters to the mining model and apply these filters during learning and

testing phases to easily generate the associated models on the data subsets, develop and compare multiple forecasting models and then take action on results.

7 References

- [1] M. El Fouki, N. Aknin, and K. E. El. Kadiri, "Intelligent Adapted e-Learning System based on Deep Reinforcement Learning," in Proceedings of the 2nd International Conference on Computing and Wireless Communication Systems - ICCWCS'17, 2017, pp. 1–6.
- [2] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, no. 1, p. 1, Dec. 2015. <https://doi.org/10.1186/s40537-014-0007-7>
- [3] G. E. Hinton, "Reducing the Dimensionality of Data with Neural Networks," *Science* (80-.), vol. 313, no. 5786, pp. 504–507, Jul. 2006.
- [4] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemom. Intell. Lab. Syst.*, vol. 2, no. 1–3, pp. 37–52, Aug. 1987. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
- [5] N. Nehiri and N. Aknin, "New Construction Process of Standardized Learner Profile," in Proceedings of the 2nd International Conference on Computing and Wireless Communication Systems - ICCWCS'17, 2017, pp. 1–5. <https://doi.org/10.1145/3167486.3167573>
- [6] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414–422, 2015. <https://doi.org/10.1016/j.procs.2015.12.157>
- [7] G. Hinton et al., "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012. <https://doi.org/10.1109/MSP.2012.2205597>
- [8] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016. <https://doi.org/10.1016/j.neucom.2015.09.116>
- [9] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequential Deep Learning for Human Action Recognition," in *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2011, pp. 29–39.
- [10] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *Int. J. Rob. Res.*, vol. 34, no. 4–5, pp. 705–724, Apr. 2015. <https://doi.org/10.1177/0278364914549607>
- [11] E. Ohn-Bar and M. M. Trivedi, "Looking at Humans in the Age of Self-Driving and Highly Automated Vehicles," *IEEE Trans. Intell. Veh.*, vol. 1, no. 1, pp. 90–104, Mar. 2016. <https://doi.org/10.1109/TIV.2016.2571067>
- [12] L. Deng, "Deep Learning: Methods and Applications," *Found. Trends® Signal Process.*, vol. 7, no. 3–4, pp. 197–387, 2014.
- [13] Y. Bengio, "Learning Deep Architectures for AI," *Found. Trends® Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009. <https://doi.org/10.1561/2200000006>
- [14] K. G. Kim, "Book Review: Deep Learning," *Healthc. Inform. Res.*, vol. 22, no. 4, p. 351, 2016. <https://doi.org/10.4258/hir.2016.22.4.351>
- [15] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Brief. Bioinform.*, p. bbw068, Jul. 2016. <https://doi.org/10.1093/bib/bbw068>

- [16] Y. Jia et al., “Caffe: Convolutional Architecture for Fast Feature Embedding,” in Proceedings of the ACM International Conference on Multimedia - MM '14, 2014, pp. 675–678. <https://doi.org/10.1145/2647868.2654889>
- [17] B. Minaei-Bidgoli, D. a. Kashy, G. Kortemeyer, and W. F. Punch, “Predicting student performance: an application of data mining methods with an educational web-based system,” 33rd Annu. Front. Educ. 2003. FIE 2003., vol. 1, p. T2A_13-T2A_18, 2003.
- [18] L. Shang, J. Cheng Lv, and Z. Yi, “Rigid medical image registration using PCA neural network,” *Neurocomputing*, vol. 69, no. 13–15, pp. 1717–1722, Aug. 2006. <https://doi.org/10.1016/j.neucom.2006.01.007>
- [19] K. Li, C. Hu, G. Liu, and W. Xue, “Building’s electricity consumption prediction using optimized artificial neural networks and principal component analysis,” *Energy Build.*, vol. 108, pp. 106–113, Dec. 2015. <https://doi.org/10.1016/j.enbuild.2015.09.002>
- [20] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, “Multi-column deep neural network for traffic sign classification,” *Neural Networks*, vol. 32, pp. 333–338, Aug. 2012. <https://doi.org/10.1016/j.neunet.2012.02.023>
- [21] J. Saxe and K. Berlin, “Deep Neural Network Based Malware Detection Using Two Dimensional Binary Program Features,” 2015.
- [22] I. T. Jolliffe and J. Cadima, “Principal component analysis: a review and recent developments,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, p. 20150202, Apr. 2016. <https://doi.org/10.1098/rsta.2015.0202>
- [23] B. Guo, R. Zhang, G. Xu, C. Shi, and L. Yang, “Predicting Students Performance in Educational Data Mining,” 2015 Int. Symp. Educ. Technol., pp. 125–128, 2015. <https://doi.org/10.1109/ISET.2015.33>
- [24] V. O. Oladokun, A. T. Adebajo, and O. E. Charles-Owaba, “Predicting students’ academic performance using artificial neural network: A case study of an engineering course,” *Pacific J. Sci. Technol.*, vol. 9, no. 1, pp. 72–79, 2008.
- [25] ASIF, Raheela, MERCERON, Agathe, ALI, Syed Abbas, et al. Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 2017, vol. 113, p. 177-194. <https://doi.org/10.1016/j.compedu.2017.05.007>
- [26] D. Kabakchieva, “Predicting student performance by using data mining methods for classification,” *Cybern. Inf. Technol.*, vol. 13, no. 1, pp. 61–72, 2013. <https://doi.org/10.2478/cait-2013-0006>
- [27] V. Ramesh, P. Parkavi, and K. Ramar, “Predicting Student Performance : A Statistical and Data Mining Approach,” *Int. J. Comput. Appl.*, vol. 63, no. 8, pp. 35–39, 2013.

8 Authors

Mohammed El Fouki received the Master degree in Computer Science in 2014 from Abdelmalek Essaadi University in Tetuan, Morocco. Currently, he is a PhD Candidate and member of Computer Science, Operational Research and Applied Statistics Laboratory in the same university.

Noura Aknin Professor of Electrical & Computer Engineering at Abdelmalek Essaadi University since 2000. She received PhD degree in Electrical Engineering in 1998. She is the Head of Research Unit Information Technology and Modeling Systems. She is the Co-founder of the IEEE Morocco Section since November 2004 and a member of several IEEE societies. Noura AKNIN is R&D project manager/member related to new technologies and their applications. She was a chair of several

conferences and she has been involved in the organizing and in the Scientific Committees of several international conferences held worldwide dealing with computer science and applications. (noura.aknin@uae.ac.ma)

Kamal Eddine El Kadiri is a Professor of computer science and mathematics at Abdelmalek Essaadi University, Morocco, since 1985. He received the “Graduate thesis” degree in Data analysis at Paris VI University in 1984 and “Thesis status” degree in Datamining and Computer Science from Granada in 1994. Currently, he is the Director of Computer Science, Operational Research and Applied Statistics Laboratory since 2007. Kamal Eddine El Kadiri is responsible of several R&D projects and member of several International Research projects. He is also part of many boards of international journals and international conferences. (elkadiri@uae.ma)

Article submitted 20 May 2018. Resubmitted 29 August 2018. Final acceptance 03 September 2018. Final version published as submitted by the authors.