

A New Approach to Predicting Learner Performance with Reduced Forgetting

Dagou Dangui Augustin Sylvain Legrand KOFFI¹, Tchimou N'TAKPE²
Assouhoun ADJE³, Souleymane OUMTANAGA⁴

Unité de Formation et de Recherche des Mathématiques et Informatique (UFR-MI)
Université Félix Houphouët-Boigny (UFHB), Abidjan, Côte d'Ivoire^{1,3}

Laboratoire de Mathématiques et Informatique, Université Nangui Abrogoua, Abidjan, Côte d'Ivoire²

Laboratoire de Recherche en Informatique et Télécommunication

Institut National Polytechnique Félix Houphouët-Boigny, Yamoussoukro, Côte d'Ivoire⁴

Abstract—The work on predicting learner performance allows researchers through machine learning methods to participate in the improvement of e-learning. This improvement allows, little by little, e-learning to be promoted and adopted by several educational structures around the world. Neural networks, widely used in various performance prediction works, have made several exploits. However, factors that are highly influential in the field of learning have not been explored in machine learning models. For this reason, our study attempts to show the importance of the forgetting factor in the learning system. Thus, to contribute to the improvement of accuracy in performance predictions. The interest being to draw the attention of researchers in this field to very influential factors that are not exploited. Our model takes into account the study of the forgetting factor in neural networks. The objective is to show the importance of attenuation of the forgetting, on the quality of performance predictions in e-learning. Our model is compared to those based on Random Forest and linear regression algorithms. The results of our study show first that neural networks (95.20%) are better than Random Forest (95.15%) and linear regression (93.80%). Then, with the attenuation of forgetting, these algorithms give 96.63%, 95.85% and 93.80% respectively. This work allowed us to show the great relevance of oblivion in neural networks. Thus, the exploration of other unexploited factors will make better performance prediction models.

Keywords—Performance prediction; e-learning; artificial neural networks; forgetting factor

I. INTRODUCTION

Most of the work on e-learning recommendation systems focuses on the construction of recommendation models. The aim of this recommendation model building work is to improve the recommendations of learning objects for learners [1]. The prediction of learner performance in education, particularly in e-learning, is one of the most studied research areas. It is part of the quest to improve recommendations in recommendation systems [2]. Predictions of performance in e-learning systems require a good knowledge of the users in order to provide better accuracy.

However, eLearning systems, despite having large amounts of information about their users, suffer from the lack of some very relevant information [3]. This information would help to improve the results of performance predictions. As a result, the search for highly relevant information is now at the heart of

research in the field of learning [3]. It is a challenge for researchers.

Our study in this paper focuses on improving the accuracy of learner performance predictions. Thus, we start from the fact that the use of relevant data in weak algorithms is less advantageous than the use of relevant data in powerful algorithms. For this, we use neural networks, one of the most efficient learning machine algorithms in the field of learner performance prediction [4] [5]. We use the forgetting factor in our study. It is a very important factor, which is widely used in the field of learning psychology [6]. Taking into account the attenuation of the forgetting factor in learners in neural networks is at the heart of our study.

The paper is organized as follows: Section 2 presents related work. Section 3 presents neural network algorithms and other algorithms for comparison. Section 4 presents our new approach. Section 5 presents an evaluation of our method and a discussion about our approach. Section 6 concludes this manuscript.

II. RELATED WORK

Neural networks techniques are used in many fields and today are widely used in e-learning because of their high efficiency. Oladokun et al [4], conducted a study to predict the performance of candidates likely to succeed at university using neural networks to improve the quality of university degrees. The results of this study show a performance of more than 70% and thus demonstrate the ability of neural networks to improve university admission systems. However, the authors point out the limits in the search for relevant information to make the model more effective. Arsad et al [7], for their part, also propose a neural network model for the prediction of student performance from the entry level to registration. Following the basic subjects that the student takes in the course, the results of the study showed that there is a direct correlation between the students' results for the first semester subjects with the final academic performance regardless of their gender. Thus, based on these results, the authors believe that a strategic study can be undertaken during study periods to improve students' final performance. Knowing the advantages of predicting educational performance, which helps in decision making to improve educational services, Chen et al [8] conducted a study, allowing them to propose a neural network model to predict

student performance in standardized examinations. Two meta-heuristic algorithms were used separately to form the feed-forward network for prediction and optimized the interlayer weights and biases of the neural network. Given the quality of the results, the model was designed to help students with admission procedures and to strengthen the system of services in educational institutions. Faced with the non-existence of methods for predicting student performance for some higher education institutions, Shahiri et al [5] conducted a study to predict student performance using data mining techniques on courses taught. Using performance prediction algorithms, the aim was to determine the most important factors in the learners' data for better planning of courses during study periods in order to improve the performance of their learning and teaching process. Among the prediction algorithms studied, the neural network (98%) shows the highest performance, followed by the decision tree (91%), the support vector machine (83%), the nearest neighbor cases (83%) and the Naive Bayes (76%). Jishan et al [9] present a study aimed at improving the accuracy of student performance prediction models. To do this, their study focused on data pre-processing using an oversampling technique, Synthetic Minority Over-Sampling (SMOTE) and a discretization method known as Optimal Equal Width Binning, and applied to three classification algorithms, the Naive Bayes, the decision tree and the neural network. The result of this study shows that the accuracy of the prediction models improves when using the two proposed pre-processing methods, with the neural network at the first rank followed by the Naive Bayes classifier. Kouser et al [10] are looking for ways to improve the results provided by the unstructured and low-level information collected, in order to make good and better predictions of student performance. To this end, they use exploratory methods for the analysis of raw data to extract high quality information. Four variables of students' daily activities on the Moodle platform were used in the construction of a neural network model for predicting good student grades for classes. Abu Zohair [11], faced with the difficulties encountered for predicting performance, which is not credible in education because of the small size of the data, conducted a study to find out if it was possible to obtain an accurate prediction rate by training and mobilizing students with a small data set by identifying key indicators in the data set. The results of this study indicate that the media vector machine and the learning of discriminant analysis algorithms give test rates of accuracy and reliability of 98.5%. In response to the problems of drop-out and delayed graduation, Umar [12] is conducting a study to predict poor student performance. These results can then be used for academic follow-up to alleviate these problems. Thus, a neural network capable of predicting a student's overall average using the student's personal data has been designed. The results were 73.68% correctly predicted performance with also an accuracy of 66.67% of the students likely to drop out or experience a delay before graduation. Raga and Raga [13] conducted an experiment to develop a model for predicting student performance in order to measure the non-linear predictive power of neural networks in a blended learning environment. But first, the hyperparameters of the model were determined by a series of experiments. The results of this experiment show, for a single course in the first month, an accuracy of 91.07%

with a ROC_AUC score of 0.88 which improves as the accumulation progresses. But for the mid-term results, the highest precision was 80.36% and the ROC_AUC score was 0.70. Azzi et al. en [14], propose an approach based on artificial neural networks in order to address the problems of customization of e-learning systems. In particular, they address the problem of designing courses based on the background of the learners. The role of the proposed model is to mimic the course designer in order to create customized courses for learners. The system is thus able to choose the appropriate content to improve the learner's performance.

The previous literature review presents neural networks that are widely used in performance prediction work in e-learning. Precisely because of the power of its algorithms, which are still proving their worth in this field.

Studies on predicting learner performance and taking into account forgetting have also been carried out. Nguyen et al [2] propose a study for the prediction of student performance. They use tensor factorization methods to implicitly take into account latent factors and temporal effect. The results of their proposed approaches show that they are promising and appropriate in improving prediction results. The authors also argue that a predictive approach could be used to account for the sequential effect. Nedungadi and Remya [15] propose a new PC-BKT model, an improved model of the existing BKT which generally sets the forgetting parameter to 0. The results provided by the PC-BKT show the percentage of classification errors reduced as the algorithm adjusts the learning rate of a skill over the duration between the last uses. Thus, they state that the time it takes a student to start forgetting the skill is 30 days.

The observation made at the end of this second literature review is the scarcity of performance prediction works taking into account forgetting. Even rarer are these works with neural networks.

Studies have also been carried out on its great influence in the field of learning. Ziegler [6] studied the process of forgetting empirically using data collected on the brain, spinal cord and nerves in an experiment in which a group of 58 elderly people took part in training courses followed by examinations. The results are such that the rate of forgetfulness was higher for half, and had no huge change after 118 days. The reason is that during this time, the students had revision times. Krondorfer [16] in his work proposes to differentiate between forgetting and remembering in order to help erode paralyzing traumatic memories. Thus, he argues that there are no impeccable distinctions or clearly marked boundaries between these terms. Remembrance is elevated to the status of unquestionable virtue and forgetting is mocked as undesirable and reprehensible. Casey and Olivera [17] set out in a study to clarify the relationship between organizational memory and forgetting, and to identify areas that need to be developed to improve the understanding of memory constructs. Thus, they argue that the dynamic nature of organizational knowledge, the role of time in the way organizations retain knowledge, and the role of power dynamics in what and how organizations choose to remember and forget could address their concern. Gordon [18] in one essay proposes to reflect on the meaning of the

tension between remembering and forgetting in the context of historical and tragic events. He argued that an ethic of remembering and forgetting could enable victims of trauma to understand not only the sources of their suffering but also to take responsibility for their own liberation. Gan and Zeng [19] are conducting a study with the aim of improving the speed of convergence of iterative learning, controlling it and reducing the fluctuation of system error. For this purpose, they use a class of steady-state linear systems with a variable forgetting factor. The results of the work give the algorithm is efficient and that the convergence speed is improved with a low error rate. Boutis et al [20] conducted a study to determine the rate of knowledge degradation over time using the forgetting curve. Thus, by a test on 106 participants measuring the degradation over time for 12 months, the conclusion drawn was that the degradation of learning was attenuated every two months. These results can thus influence the scheduling of refresher courses.

This third review of the literature presents the study on forgetting in more research work. Which shows its great importance.

III. MACHINE LEARNING ALGORITHMS

A. Artificial Neural Networks (ANN)

Neural networks are very powerful algorithms used in Artificial Intelligence, especially in the field of machine learning. It has been designed to approach problem solving in the same way as the human brain does. It has the ability to solve problems of great complexity [10]. The architecture of neural networks consists of three types of layers, the hidden layers located in between, the input layer, and the output layer. Each layer may consist of at least one node called a neuron. The neurons of the different layers are connected to each other by synaptic weights as shown in Figure 1. Neural networks vary according to the type of configuration. The simplest of these configurations is the perceptron.

Data flows through neural networks from input to output through a process called forward propagation. In this process, neuron outputs are determined by the arithmetic operation of applying a neuron activation function to the sum of all inputs. The activation function is a stimulation threshold which, when reached, causes the neuron to respond. Figure 2 shows a representation of a neuron. There are several types of activation functions, which can be both linear and non-linear depending on the objectives to be achieved. The most commonly used is the sigmoid or logistic function. However, other more powerful activation functions such as the reread [21] exist.

After forward propagation, the outputs and errors of the neural network outputs are determined. Back propagation is performed to adjust the values of the neural network weights randomly initialized at the beginning.

B. Random Forest

The Random Forest is also an algorithm in the field of machine learning. It makes it possible to make a classification of whole. It is more effective in the predictions for sets than the particular predictions [22]. Studies have shown that the Random Forest is based on the most powerful nonparametric

classifiers [23] [24]. In this logic, this classifier is adapted to our study for a good measurement of our model. It is based on the principle of decision trees. As its name indicates, it is composed of a large number of decision trees which function as a set. Each individual tree in Random Forest makes a class prediction. Thus, the class with the most votes becomes the class considered for the Random Forest prediction.

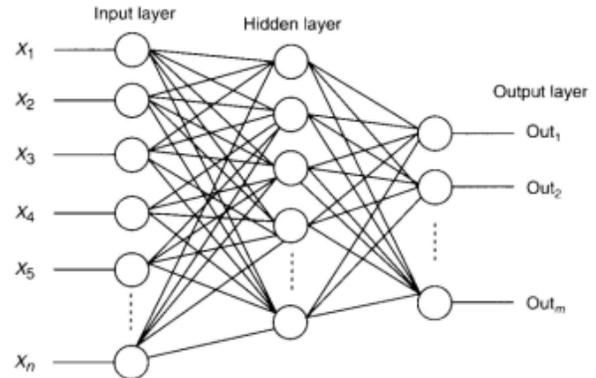


Fig. 1. Neural Network Architecture [10].

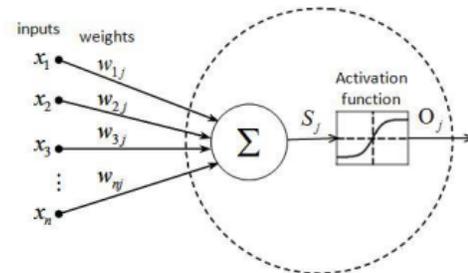


Fig. 2. Representation of a Neuron [10].

C. Linear Regression

Linear regression is also a self-learning algorithm [25]. It is an approach for modelling the relationship between a dependent variable and one or more explicit variables. The objective of this algorithm is to determine the hyperparameters of the linear model, formed from a set of data. It allows the design of a rectilinear or curvilinear function that best approximates the elements of the data set. Linear regression is one of the best known and most widely used methods in statistics for the analysis of quantitative data.

IV. NEW APPROACH

Many performance predictions works have been performed by neural networks because of their high efficiency [5] [9]. The approach proposed in this work is also based on neural networks. It integrates a very important latent factor which is oblivion. This is a very influential factor in the field of learning. Moreover, performance prediction works including this factor are rare in the literature. Even more so with neural networks in predicting learner performance. The objective is to show the impact that this factor has on performance predictions in neural networks. The history of learner performance is taken into account, as well as other general factors at the input of the network. The goal is to bring more precision in performance predictions. (Ebbinghaus, H. (1885/1962)), one of the fathers

of experimental psychology who showed by the results of his experiments on the experimental study of memory and the learning process, that the forgetting curve describes a decreasing exponential form [26] [27], as shown in figure 3.

This Ebbinghaus theory is still relevant today and is widely used by many researchers working on the cognitive. It tends to show that the rate of information loss by the human brain is exponential. The work of L. Averell and A. Heathcote in 2011 [29], focused on determining the mathematical form of forgetting. They proposed three candidate functions that best express forgetting, namely the exponential $e^{-\alpha t}$, the pareto $(1 + \gamma t)^{-\beta}$ and the power $(1 + t)^{-\beta}$. The exponential function is the most appreciated for the description of forgetting compared to the other two functions. Their research confirms Ebbinghaus' theory. The proposal of our model for managing forgetting in e-learning is based on the exponential form of forgetting defined in the following equation,

$$R = e^{-\alpha t} \quad (1)$$

where R is the memory retention and α is the forgetting rate.

The forgetting rate, on the other hand, is a function of the memory strength F, as presented in the following equation.

$$\alpha = \frac{1}{F} \quad (2)$$

[30] Thus, the greater the memory strength, the lower the forgetfulness rate, which increases in (1) memory retention R, thus reducing forgetfulness. Similarly, the smaller the memory strength, the higher the forgetting rate, which normally reduces memory retention, thus increasing forgetting. According to the work of Ebbinghaus, memory depends essentially on the number of repetitions of a learning element and the time spent reviewing what has been learned [27]. From this reflection, memory is strengthened with the high number of repetitions of the learner. Thus, to reduce forgetting in our model, the number of repetitions of learning elements must be increased. However, the set of data available to us for our work does not include a variable signifying the number of repetitions for the learners' learning. However, our data set has a variable that means absences from classes by learners. Repetitions and absences have opposite effects on learning. Thus, a learner's retention capacity increases as the number of repetitions increases. It also increases when the number of absences from classes decreases. However, the two increases are not the same in both cases. To mitigate forgetting in our model, we consider the rate of absences at courses instead of the number of repetitions of learning elements. Thus, by reducing the content of the absence variable, retention capacity increases. However, it is better to take into account the number of repetitions if it is included such as a variable in a dataset for a study. Therefore, the manipulation of the absence rate, free time after school and the learner's study time would be tantamount to acting on forgetting so as to see its impact on the prediction of performance. Thus, likely to lead to good decisions for homework sessions in order to optimize academic performance.

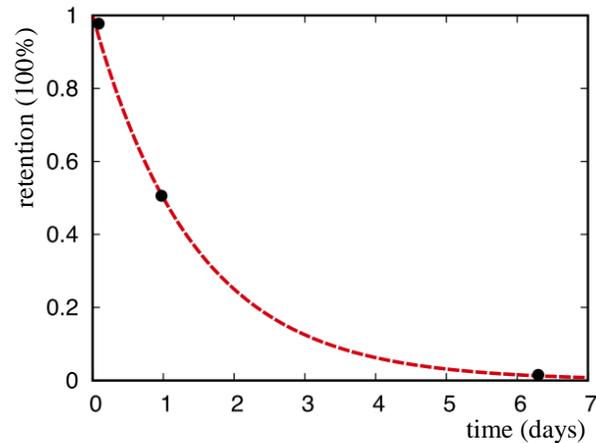


Fig. 3. The Ebbinghaus Forgetting Curve.[28].

Our experiments are done in two steps. In the first step, a study of learner performance predictions is made by designing three prediction models. The first model is done using neural networks, the second using Random Forest and the third using linear regression. A comparison of the performance results obtained from our three models is then performed. The objective of this first step is to confirm or refute the results of the work on the performance of neural networks as the best predictor of e-learning. Assertion made in the literature review.

In the second step, the experience of the first step is experienced again, this time taking into account the reduction of forgetting. To this end, reductions in the rate of absence and free time are made. These are factors on which forgetting depends. The reduction of each of these variables aims to reduce forgetfulness among learners. Thus, for this phase, the objective is to determine approximately the probable performance of the learners with reduced forgetting. The results of the predictions with reduced forgetting will be compared with the results of the experiments in the first stage. Thus, it can be shown the contribution of the attenuation of the forgetting in the predictions of performance of the learners.

The reductions of learners' absences at class and free time, in our experience, are respectively 90% of the content of the variable "Absences" and 20% of the content of the variable "Freetime" of the data set. As for the learner study time variable, it remains unchanged. Ideally, it should be increased in order to intensify the reduction of forgetting. The reduction rates proposed for the absence rate and free time are not based on any theory. They are taken randomly with the sole aim of reducing forgetfulness.

A. Data Set

The dataset used in this work is available in the UCI's database. It consists of 395 samples with 30 variables. It also includes three data sets G1, G2 and G3 which are the successive results of evaluations. Each of these variables is a whole score between 0 and 20. The objective of our study will be to predict the G3 assessment scores of learners assumed to be unknown and assuming that G1 and G2 scores are already made. There is no missing value data in this set. The following list gives a description of the 30 variables.

1 school	student's school (binary: "GP" - Gabriel Pereira or "MS" - Mousinho da Silveira)
2 sex	student's sex (binary: "F" - female or "M" - male)
3 age	student's age (numeric: from 15 to 22)
4 address	student's home address type (binary: "U" - urban or "R" - rural)
5 famsize	family size (binary: "LE3" - less or equal to 3 or "GT3" - greater than 3)
6 Pstatus	parent's cohabitation status (binary: "T" - living together or "A" - apart)
7 Medu	mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
8 Fedu	father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
9 Mjob	mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")
10 Fjob	father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")
11 reason	reason to choose this school (nominal: close to "home", school "reputation", "course" preference or "other")
12 guardian	student's guardian (nominal: "mother", "father" or "other")
13 traveltime	home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
14 studytime	weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15 failures	number of past class failures (numeric: n if 1<=n<3, else 4)
16 schoolsup	extra educational support (binary: yes or no)
17 famsup	family educational support (binary: yes or no)
18 paid	extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19 activities	extra-curricular activities (binary: yes or no)
20 nursery	attended nursery school (binary: yes or no)
21 higher	wants to take higher education (binary: yes or no)
22 internet	Internet access at home (binary: yes or no)
23 romantic	with a romantic relationship (binary: yes or no)
24 famrel	quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25 freetime	free time after school (numeric: from 1 - very low to 5 - very high)
26 goout	going out with friends (numeric: from 1 - very

low to 5 - very high)

27 Dalc	workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28 Walc	weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29 health	current health status (numeric: from 1 - very bad to 5 - very good)
30 absences	number of school absences (numeric: from 0 to 93)

Evaluation grades:

31 G1	first period grade (numeric: from 0 to 20)
32 G2	second period grade (numeric: from 0 to 20)
33 G3	final grade (numeric: from 0 to 20, output target)

Dataset source:

(<https://archive.ics.uci.edu/ml/datasets/Student+Performance>)

B. Splitting the Data Set

Splitting the data set in the machine learning process consists of dividing the data used into two parts. The first part, which will be the largest portion, is reserved for the training phase of the model. It is usually greater than or equal to 60%. The second part, which is obviously the smaller one, is dedicated to the test phase of the trained model. For the experimentation of our model, the data set was divided into two parts. One part for training with a proportion of 70% and the other part for testing with a proportion of 30%.

C. Configuration of our Neural Network

Our proposed neural network model is shown in Figure 4 and is configured on three layers, one input layer, another as a hidden layer and the third is the output layer. The input layer has 32 neurons for the 30 variables of the dataset and G1 and G2 notes to take into account the learners' prior skills. The hidden layer also has 32 neurons. The output layer has a single neuron. The neuron activation function used is the "relu" because it is more efficient than the sigmoid [21], with "adam" as solver and a constant learning rate, initialized at 0.01. the maximum number of iterations is 500.

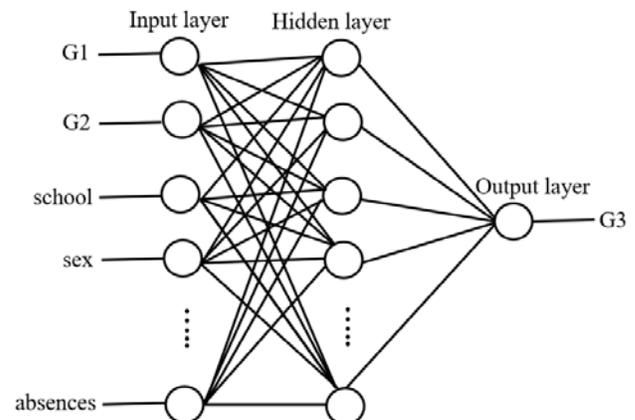


Fig. 4. Our Neural Network Model.

V. EXPERIMENTAL RESULTS AND DISCUSSION

Our experiments were carried out with a four-core I5 computer with a processor speed ranging from 1.70 to 2.40. The computer was equipped with an I5 processor with four cores. This computer works with a 12 GB RAM memory. The programming language used is the python language. After the experiment follows the evaluation phase. In this phase, the root means square error and the confusion matrix are used for the evaluation of the experiments in this study. The error is thus determined by the following formula:

$$RMSE = \sqrt{\frac{\sum_{ui}(r_{ui}-\hat{r}_{ui})^2}{n}} \quad (3)$$

The confusion matrix defines the metrics accuracy, precision, recall and F-measure as shown in equations (4), (5), (6) and (7) respectively. These equations are a function of the information in Table 1, defined as follows:

TABLE I. CONFUSION MATRICES

		Detected	
		Positive	Negative
Current	Positive	True positives (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

TP : Positive class result correctly predicted;

TN: negative class result: negative class result predicted correctly;

FP : result of the incorrectly predicted positive class;

FN : negative class result: result of the negative class incorrectly predicted.

The confusion matrix thus defines the following metrics:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F - Measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (7)$$

Accuracy: refers to the proportion of correct predictions;

Precision: refers to the proportion of correct predictions among positive predictions;

Recall: refers to the proportion of positives that are correctly identified;

F-Measure: Used to evaluate a compromise between recall and precision.

A. Results and Discussion of Step 1 of the Experiment

This step consists of comparing the performance of the algorithms used in this study. The goal is to confirm the claims of the literature review indicating neural networks as the best predictor of learner performance in e-learning. Table 2 presents

the results of the confusion matrix metrics for the artificial neural network algorithms, Random forest and Linear regression. Table 3 presents the results of the errors obtained from these algorithms during the experiment.

At the end of these first phase experiments, the artificial neural networks show the best results for each of the accuracy, precision, recall and F-Measure metrics of 94.94%, 95.20%, 97.36% and 96.27% respectively. The error rate is 0.2250. The performance of the neural networks is followed by the performance of the Random Forest with 94.17 %, 95.15 %, 96.23 % and 95.69 % for the accuracy, precision, recall and F-Measure metrics respectively. The error rate for Random Forest is 0.2413. Linear regression comes last with 92.15%, 93.80%, 94.32% and 93.98%, in the same order of accuracy, precision, recall and F-Measure metrics. The error rate is 0.2541. The results of this first phase confirm the performance of neural networks as the best predictors of e-learning performance as indicated in the literature review. On the other hand, the enormous performance of Random Forest should be underlined. Could it be more effective than neural networks under other conditions and circumstances, on other types and scales of data? The question remains for possible studies.

B. Results and Discussion of Step 2 of the Experiment

This step consists of determining the performance of the algorithms used in this study by reducing the effects of forgetting among learners. The performances obtained are compared to the performances obtained during the first step according to the algorithms. Table 4 presents the results of the confusion matrix metrics for the artificial neural network, Random forest and Linear regression algorithms. Table 5 presents the results of the errors obtained from these algorithms during the experiment.

TABLE II. COMPARISON RESULTS OF ALGORITHMS WITH OMISSION

Algorithms	Accuracy	Precision	Recall	F- Measure
Artificial Neural Network	94.94 %	95.20 %	97.36 %	96.27 %
Random forest	94.17 %	95.15 %	96.23 %	95.69 %
Linear regression	92.15 %	93.80 %	94.32 %	93.98 %

TABLE III. ALGORITHM ERRORS WITH OVERLOOKED ALGORITHMS

Algorithms	RMSE
Artificial Neural Network	0,2250
Random forest	0,2413
Linear regression	0,2541

TABLE IV. COMPARISON RESULTS OF ALGORITHMS WITH ATTENUATED FORGETTING

Algorithms	Accuracy	Precision	Recall	F- Measure
Artificial Neural Network	95.95 %	96.63 %	97.36 %	97 %
Random forest	94.43 %	95.85 %	95.85 %	95.85 %
Linear regression	92.15 %	93.80 %	94.32 %	93.98 %

TABLE V. ALGORITHM ERRORS WITH SOFT OVERLOOKED ALGORITHMS

Algorithms	RMSE
Artificial Neural Network	0,2012
Random forest	0,2360
Linear regression	0,2541

At the end of these second phase experiments, the artificial neural networks still show the best results for each of the accuracy, precision, recall and F-Measure metrics of 95.95%, 96.63%, 97.36% and 97% respectively. The error rate is 0.2012. Neural network performance is always followed by Random Forest performance with this time 94.43%, 95.85%, 95.85% and 95.85% for the accuracy, precision, recall and F-Measure metrics respectively. The error rate for the Random forest now increases to 0.2360. Linear regression comes last, but maintains the same performance as in Phase 1 with 92.15%, 93.80%, 94.32% and 93.98%, still in the same order as the accuracy, precision, recall and F-Measure metrics. The error rate is 0.2541. The results of this second phase still confirm the performance of neural networks as the best predictors of e-learning performance as indicated in the literature review, despite the reduction in forgetting.

The finding in this second phase is that there is an improvement in accuracy and precision for neural networks of 1.01% and 1.43% respectively when forgetting is reduced. For Random forest, there is an improvement in accuracy and precision of 0.26% and 0.7% respectively. However, there is no variation in the linear regression. These variations in accuracy and precision for these algorithms have two meanings:

1) Respectively, with respect to accuracy and precision, neural networks have a better ratio of correct predictions and also a better ratio of confirmations of correct positive predictions compared to the other two algorithms.

2) A greater improvement in the ratio of correct predictions and the ratio of confirmations of correct positive predictions compared to the other two algorithms.

As far as recall is concerned, there is no improvement when reducing forgetfulness for neural networks and linear regression. However, there is a regression of -0.38% for Random forest. This means that compared to Random forest, neural networks and linear regression best predict positive cases.

For F-measure, with the reduction of forgetfulness, neural networks, Random forest and linear regression increase by 0.74%, 0.16% and 0% respectively. Thus, neural networks and Random forest reduce the trade-off between recall and accuracy. This result not only confirms the better performance of neural networks compared to other algorithms but also shows an improvement in the quality of neural networks when reducing forgetfulness.

Finally, the error margins were reduced to 0, 0.0053 and 0.0238 respectively for linear regression, Random forest and neural networks with the reduction of forgetting among learners. The neural networks still show their performance with more error reduction.

VI. CONCLUSION

The objective of our study was to clarify the contribution of taking into account the forgetting factor in models for predicting learner performance in e-learning. During our work, we integrated it into an artificial neural network to create a model in order to see the impact of this factor on performance predictions in these neural networks. In this study, two other classifiers, namely Random Forest and linear regression, were used in order to compare their results with those of our model. At the end of this work, we observe the performance of neural networks. With the process of reducing forgetfulness proposed by our model, neural networks give even better results. This will allow us to better specify the performance of students in order to adopt better strategies in decisions about their training. Our theory also improves the performance of the Random Forest.

In our future work, we will extend this study to a larger scale to see what impact our theory will have on large datasets. We will also rule on the study of stress, another latent factor that has a strong influence on memory in the learning system. We will also conduct another study to identify the most important variables in a list of variables in a data set. This may help to further improve predictions of learner performance.

ACKNOWLEDGMENT

At the Unité de Formation et de Recherche des Mathématiques et Informatique (UFR-MI) of the Université Félix Houphouët-Boigny (UFHB) and the Laboratoire de Recherche en Informatique et Télécommunication (LARIT). Also, to the UCI for its database.

REFERENCES

- [1] K. I. Ghauth et N. A. Abdullah, "Learning materials recommendation using good learners' ratings and content-based filtering", *Education Tech Research Dev*, vol. 58, n° 6, p. 711-727, déc. 2010, doi: 10.1007/s11423-010-9155-4.
- [2] N. Thai-nghe, L. Drumond, R. Nanopoulos, et L. Schmidt-thieme, "Matrix and tensor factorization for predicting student performance", in *In Proceedings of the 3rd International Conference on Computer Supported Education (CSEDU)*, 2011.
- [3] H. Al-Samarraie, B. K. Teng, A. I. Alzahrani, et N. Alalwan, "E-learning continuance satisfaction in higher education: a unified perspective from instructors and students", *Studies in Higher Education*, vol. 43, n° 11, p. 2003-2019, nov. 2018, doi: 10.1080/03075079.2017.1298088.
- [4] V. Oladokun, A. Adebajo, B. Sc, et O. Charles-Owaba, "Predicting Students Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course", *The Pacific Journal of Science and Technology*, vol. 9, April 1, 2.
- [5] A. M. Shahiri, W. Husain, et N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques", *Procedia Computer Science*, vol. 72, p. 414-422, janv. 2015, doi: 10.1016/j.procs.2015.12.157.
- [6] L. H. Ziegler, "Learning and Forgetting", *The Journal of Higher Education*, vol. 2, n° 3, p. 144-146, mars 1931, doi: 10.1080/00221546.1931.11772362.
- [7] P. M. Arsad, N. Buniyamin, et J. A. Manan, "Neural network model to predict electrical students' academic performance", in *2012 4th International Congress on Engineering Education*, déc. 2012, p. 1-5, doi: 10.1109/ICEED.2012.6779270.
- [8] J.-F. Chen, H.-N. Hsieh, et Q. H. Do, "Predicting Student Academic Performance: A Comparison of Two Meta-Heuristic Algorithms Inspired by Cuckoo Birds for Training Neural Networks", *Algorithms*, vol. 7, n° 4, p. 538-553, déc. 2014, doi: 10.3390/a7040538.

- [9] S. Jishan, R. Rashu, N. Haque, et R. M. Rahman, "Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique", *Decision Analytics*, vol. 2, n° 1, p. 1- 25, 2015, doi: 10.1186/s40165-014-0010-2.
- [10] R. Kouser, J. D. Joann, et K. Suganya, "Forecasting Student Academic Performance by Decision Tree Learning using Artificial Neural Networks", 2017. <https://www.semanticscholar.org/paper/FORECASTING-STUDENT-ACADEMIC-PERFORMANCE-BY-TREE-Kouser-Joann/5b116400ef849d120602656719888f5c2e3ec0e2>.
- [11] L. M. Abu Zohair, "Prediction of Student's performance by modelling small dataset size", *International Journal of Educational Technology in Higher Education*, vol. 16, n° 1, p. 27, août 2019, doi: 10.1186/s41239-019-0160-3.
- [12] M. A. Umar, "Student Academic Performance Prediction using Artificial Neural Networks: A Case Study", *International Journal of Computer Applications*, vol. 178, p. 24- 29, sept. 2019, doi: 10.5120/ijca201919387.
- [13] R. C. Raga et J. D. Raga, "Early Prediction of Student Performance in Blended Learning Courses Using Deep Neural Networks", in *2019 International Symposium on Educational Technology (ISET)*, juill. 2019, p. 39- 43, doi: 10.1109/ISET.2019.00018.
- [14] I. Azzi, A. Jeghal, A. Radouane, A. Yahyaouy, et H. Tairi, « Approach Based on Artificial Neural Network to Improve Personalization in Adaptive E-Learning Systems », in *Embedded Systems and Artificial Intelligence*, Singapore, 2020, p. 463- 474, doi: 10.1007/978-981-15-0947-6_44.
- [15] P. Nedungadi et M. S. Remya, "Incorporating forgetting in the Personalized, Clustered, Bayesian Knowledge Tracing (PC-BKT) model", in *2015 International Conference on Cognitive Computing and Information Processing(CCIP)*, mars 2015, p. 1- 5, doi: 10.1109/CCIP.2015.7100688.
- [16] B. Krondorfer, "IS Forgetting Reprehensible? Holocaust Remembrance and the Task of Oblivion", *Journal of Religious Ethics*, vol. 36, n° 2, p. 233- 267, juin 2008, doi: 10.1111/j.1467-9795.2008.00345.x.
- [17] A. J. Casey et F. Olivera, "Reflections on Organizational Memory and Forgetting", *Journal of Management Inquiry*, vol. 20, n° 3, p. 305- 310, sept. 2011, doi: 10.1177/1056492611408264.
- [18] M. Gordon, "Between Remembering and Forgetting", *Stud Philos Educ*, vol. 34, n° 5, p. 489- 503, sept. 2015, doi: 10.1007/s11217-014-9451-2.
- [19] Y. Gan et Q. Zeng, "Variable Gain Iterative Learning Control with Forgetting Factor", *IOP Conference Series: Materials Science and Engineering*, vol. 394, p. 052082, août 2018, doi: 10.1088/1757-899X/394/5/052082.
- [20] K. Boutis et al., "The effect of testing and feedback on the forgetting curves for radiograph interpretation skills", *Medical Teacher*, vol. 41, n° 7, p. 756- 764, juill. 2019, doi: 10.1080/0142159X.2019.1570098.
- [21] D. Pedamonti, "Comparison of non-linear activation functions for deep neural networks on MNIST classification task", avr. 2018.
- [22] M. Pal, "Random forest classifier for remote sensing classification", *International Journal of Remote Sensing*, vol. 26, n° 1, p. 217- 222, janv. 2005, doi: 10.1080/01431160412331269698.
- [23] T. Han, D. Jiang, Q. Zhao, L. Wang, et K. Yin, "Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery", *Transactions of the Institute of Measurement and Control*, vol. 40, p. 2681 - 2693, mai 2018, doi: 10.1177/0142331217708242.
- [24] M. Liu, M. Wang, J. Wang, et D. Li, "Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar", *Sensors and Actuators B: Chemical*, vol. 177, p. 970-980, févr. 2013, doi: 10.1016/j.snb.2012.11.071.
- [25] I. Naseem, R. Togneri, et M. Bennamoun, "Linear Regression for Face Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, n° 11, p. 2106- 2112, nov. 2010, doi: 10.1109/TPAMI.2010.128.
- [26] "Ebbinghaus, H. (1885/1962). *Memory: A contribution to experimental psychology*. New York: Dover".
- [27] G. A. Kimble et M. Wertheimer, *Portraits of Pioneers in Psychology*. Psychology Press, 2014.
- [28] B. Settles et B. Meeder, "A Trainable Spaced Repetition Model for Language Learning", in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany, août 2016, p. 1848-1858, doi: 10.18653/v1/P16-1174.
- [29] L. Averell et A. Heathcote, "The form of the forgetting curve and the fate of memories", *Journal of Mathematical Psychology*, vol. 55, n° 1, p. 25- 35, févr. 2011, doi: 10.1016/j.jmp.2010.08.009.
- [30] S. G. Hu et al., "Emulating the Ebbinghaus forgetting curve of the human brain with a NiO-based memristor", *Appl. Phys. Lett.*, vol. 103, n° 13, p. 133701, sept. 2013, doi: 10.1063/1.4822124.