




# XGBoost and Deep Neural Network Comparison: The Case of Teams' Performance

Filippos Giannakas<sup>(✉)</sup>, Christos Troussas, Akrivi Krouska,  
Cleo Sgouropoulou, and Ioannis Voyiatzis

Department of Informatics and Computer Engineering, University of West Attica,  
Egaleo, P. O. 12243, Athens, Greece

{fgiannakas,ctrouss,akrouska,csgouro,voyageri}@uniwa.gr

**Abstract.** In the educational setting, working in teams is considered an essential collaborative activity where various biases exist that influence the prediction of teams performance. To tackle this issue, machine learning algorithms can be properly explored and utilized. In this context, the main objective of the current paper is to explore the ability of the eXtreme Gradient Boosting (XGBoost) algorithm and a Deep Neural Network (DNN) with 4 hidden layers to make predictions about the teams' performance. The major finding of the current paper is that shallow machine learning performed better learning and prediction results than the DNN. Specifically, the XGBoost learning accuracy was found to be 100% during teams learning and production phase, while its prediction accuracy was found to be 95.60% and 93.08%, respectively for the same phases. Similarly, the learning accuracy of the DNN was found to be 89.26% and 81.23%, while its prediction accuracy was found to be 80.50% and 77.36%, during the two phases.

**Keywords:** Deep Neural Network · Machine learning · Comparison · XGBoost · Adamax · Team performance

## 1 Introduction

Nowadays, Machine Learning (ML) gains applicability in various domains where different applications were developed for addressing e.g., sentiment analysis, image recognition, natural language processing, speech recognition, and others. Especially in the educational setting, various ML applications were developed and incorporated on Intelligent Tutoring Systems (ITS), that among others, support and increase learners' engagement, retention, motivation, towards improving their learning outcomes [12–14].

The learning process is becoming more challenging when learners are engaged in a team-based learning experience. This is because collaboration and communication activities among participants seem to significantly influence students'

learning development [9]. In this context, the knowledge about the team performance is considered an essential task that reflects teams' efficiency, expectations, and learning outcomes of their members [6].

However, the literature shows poor research efforts in predicting team performance in terms of quality [1]. Instead, the research community mainly focused on exploring participants' individual characteristics and analyze the factors that influence their performance, aiming at delivering punctual interventions, avoiding to interpret how the qualities that exist during team-working experience influence teams' performance as a whole. For instance, the work of [4] presents a ubiquitous guide-learning system for tracing and enhancing students' performance by observing and evaluating their active participation. The same situation was found in the work of [8], where learners' communication activity was observed for measuring learners' performance. In the same context, the authors in [2, 15] assessed the interaction activity of Computer-Supported Collaborative Learning (CSCL) environments for concluding about students' learning performance [2, 15].

Exploring the ability of ML to make predictions about the teams' performance, the literature has shown that this is at an initial stage. The only related work found in the literature is that of [11] and [10], where the Random Forest (RF) algorithm for assessing and predicting students' learning effectiveness. Specifically, in the work of [10] the learning accuracy of their ML model was found to be 90%, using little input collaborative data. Later, the same authors in [11] trained the same ML model on a bigger scale of input data and its learning accuracy was measured at 70%. Also, the precision accuracy of their model was also calculated and found to be 0.54% and 0.61% during the learning and production phase of the teams, while the recall metric was 0.76% and 0.82%. Another research work, but for predicting the performance of team's member is that of [16]. In this work a shallow ANN with one hidden layer was enabled to predict students' academic performance in two courses. The model was trained using little input samples and the prediction accuracy was found to be 98.3%, missing other comparison results.

Concluding, while many researchers explored various machine learning algorithms for making predictions, there are missing works for exploring the ability and compare the performance of a ML with a DNN model for making predictions about the teams' performance. In this context, the major contribution of the current paper is to explore the use of the XGBoost supervised gradient boosting decision tree and a gradient descent Deep Neural Network (DNN) with four hidden layers that enables Adamax optimizer model, in terms of learning and accuracy. Summarizing, the current paper intends to answer the following two research questions:

- RQ1: Can a binary classification model being shaped and used by a ML and a DNN model for predicting the teams' performance?
- RQ2: Which is the performance of the XGBoost and DNN model?

The rest of the paper is structured as follows. The next section presents the research design by discussing the structure of the models, the dataset used for

the training and prediction phase, as well as the evaluation of the models. The results and the discussion are summarized in Sect. 3. The last section concludes the paper and provides directions to future work.

## 2 Research Design

In order to be able to answer RQ1, working in teams and their performance is formulated and explored in terms of a supervised binary classification problem. This formation will further assist the evaluation and comparison of the XGBoost and DNN model as presented below.

### 2.1 XGBoost Model

Briefly, XGBoost is a gradient boosted decision trees algorithm [3], that enables advanced regularization ( $L1$  &  $L2$ ) for assisting model generalization and control overfitting (model learn from noise and inaccurate data), for producing better performance results. In general, gradient boosting machine learning technique produces an ensemble of weak prediction models in the form of typical decision trees. A decision tree model is structured by partitioning a set of features  $X$  into a set of  $T$  non-overlapping regions (nodes) ( $R1$  to  $RT$ ). Then, a prediction is generated for each region by calculating a simple constant model  $f$  as shown in Eq. 1.

$$f(x) = \sum_{j=1}^T w_j * I(x \in R_j) \quad (1)$$

### 2.2 DNN Model

The proposed DNN structure consists of 84 input neurons, whereas the four hidden layers have 64, 32, 16, and 8 neurons, respectively. The total parameters of the model are 8.193, that all considered trainable. The linear activation function of ReLu (Rectified Linear Unit) was used in the first layers, whereas for the output layer the Sigmoid (Uni-Polar) function was selected. Also, the “Adamax” adaptive learning rate algorithm for weights optimization [7], and the “Binary Cross-Entropy” loss function were enabled. The proposed structure of the DNN model was selected after conducting a test-based evaluation of the model’s performance by varying the layers and/or the nodes of its structure.

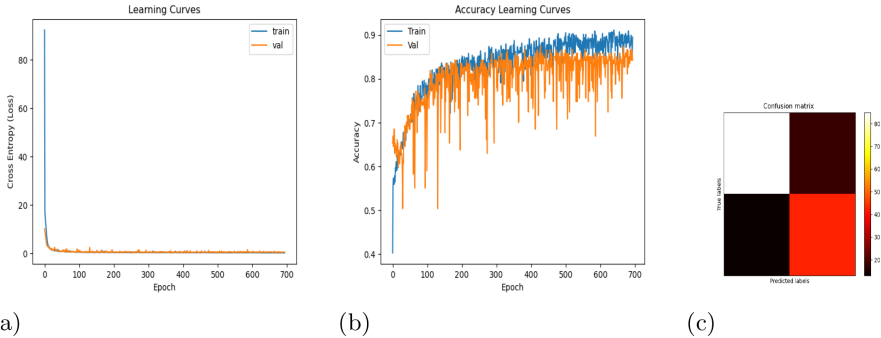
### 2.3 Evaluation Procedure

In order to evaluate both the XGBoost and the DNN models, different metrics that of the “confusion matrix” (row 1: TN-FP, row 2: FN-TP), and its related metrics for assessing accuracy, recall, and precision, were computed.

Additionally, “F1 score” and “AUC-ROC (Area Under The Curve - Receiver Operating Characteristics)” metrics were also calculated in order to compute the preciseness and the robustness of the model.

**Table 1.** Evaluation results

Model	Data	Learning	Prediction	Precision	Recall	F1 score	ROC-AUC
		Acc. (%)	Acc. (%)	Acc.			
XGBoost	Process	100	95.60	0.9778	0.8800	0.9220	0.9354
XGBoost	Product	100	93.08	0.8906	0.9344	0.9120	0.9315
DNN	Process	89.26	80.50	0.7049	0.7679	0.7350	0.8776
DNN	Product	81.23	77.36	0.7039	0.8451	0.7692	0.8558



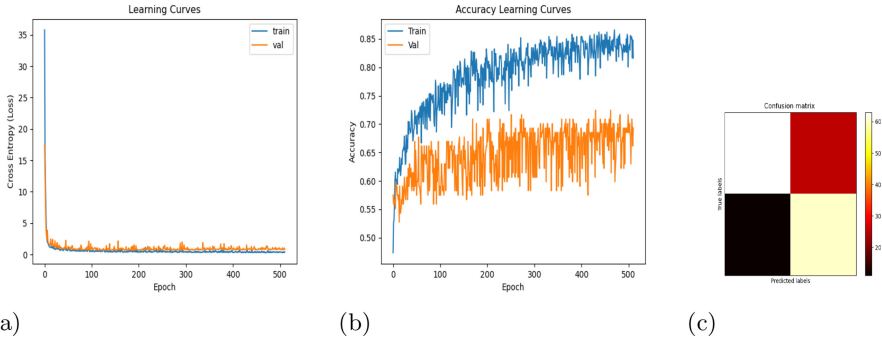
**Fig. 1.** DNN’s performance curves for the process data: (a) Cross-entropy (loss); (b) Learning accuracy; (c) Confusion matrix.

Both models were trained and evaluated using the same dataset of “Data for Software Engineering Teamwork Assessment in Education Setting Data Set (SETAP)”. The dataset contains records from different teams that attended the same software engineering course at San Francisco State University (USA), Fulda University (Germany), and Florida Atlantic University (USA), from 2012 to 2015 [11].

The SETAP project contains 30000 entries separated into two different phases that of process and product. During the process phase the learners educated how effectively they apply software engineering processes in a teamwork setting. On the contrary, in the product phase teams follow specific IT requirements for developing a software. The last column of the dataset classifies the teams’ performance into two classes [10].

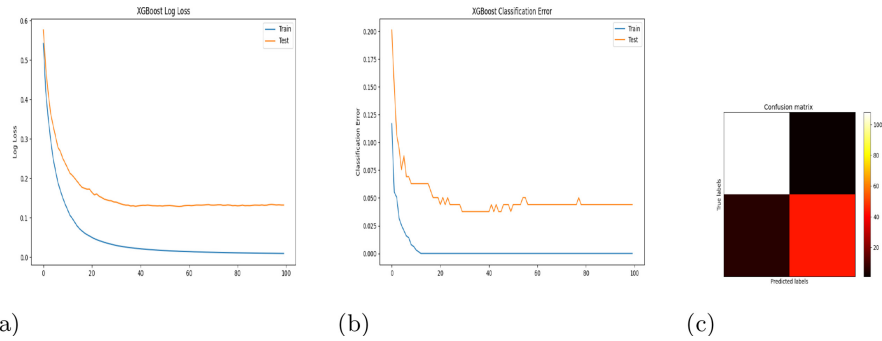
### 3 Models’ Performance Results and Discussion

In order to answer the RQ2, performance analysis was conducted for both models by computing the relevant metrics, as discussed previously in Subsect. 2.3. Specifically, the evaluation results are summarized in Table 1. Observing the table, XGBoost algorithm succeeded better accuracy results compared to the DNN model. Specifically, it was found that when the XGBoost algorithm is used,



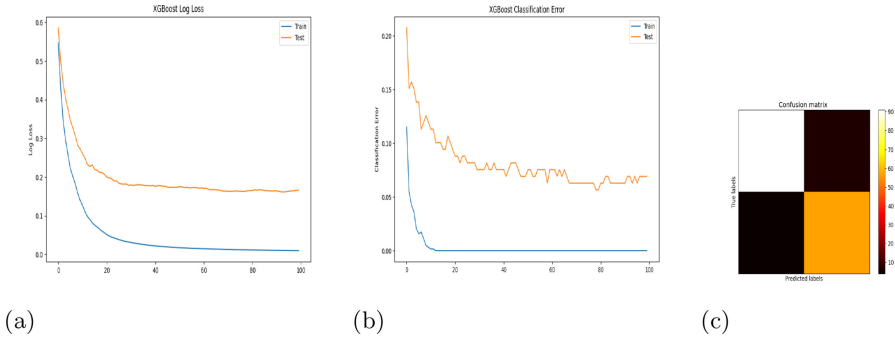
**Fig. 2.** DNN’s performance curves for the product data: (a) Cross-entropy (loss); (b) Learning accuracy; (c) Confusion matrix.

the learning and prediction accuracy was found to be at 100% and 95.60% during the process phase, whereas during the production phase the results were found to be 100% and 93.08%. Further, the precision, recall, F1 score, and ROC-AUC were found to be 0.9778, 0.8800, 0.9263, and 0.9354 for the process phase, while during the production phase the results for the same metrics were computed at 0.8906, 0.9344, 0.9120, and 0.9315.



**Fig. 3.** XGBoost’s performance curves for the process data: (a) Loss; (b) Classification error; (c) Confusion matrix.

Similarly, observing the result for the DNN model, when the data from the process phase were used, its learning and prediction accuracy was found to be at 89.26% and 80.50%, respectively. The rest of the metrics that of precision, recall, F1 score, and ROC-AUC were found to be 0.7049, 0.7679, 0.7350, and 0.8776. Same, for the product data, the DNN model succeeded 81.23% and 77.36% learning and prediction accuracy, whereas the rest of the metrics were computed at 0.7059, 0.8451, 0.7692, and 0.8558.



**Fig. 4.** XGBoost’s performance curves for the product data: (a) Loss; (b) Classification error; (c) Confusion matrix.

Further, XGBoost loss error, classification error, and confusion matrix during the process and product phases are shown in Figs. 3 and 4. Also, DNN model’s cross-entropy(loss), learning accuracy and confusion matrix curves, for the two phases are shown in Figs. 2 and 1.

To sum up, for shaping knowledge about the teams’ performance in terms of binary classification, the shallow machine learning model that enables the XGBoost algorithm performed outstanding performance results compared to the DNN model. However, we can not safely argue that the proposed ML model that enables the XGBoost algorithm is better than the DNN one since the performance of the models maybe also influenced by the quality of the input data [5].

## 4 Conclusion

The paper explores and compares the performance of the XGBoost and the DNN model that enables Adamax optimizer and Binary Cross-Entropy loss function with four hidden layers were explored and evaluated for predicting the teams’ performance. The results showed that the XGboost model outperformed the DNN by succeeding 100% learning accuracy, during the process and production phase of the teams, while its prediction accuracy was found to be 95.60% and 93.08%, for same phases. The overall learning performance of the DNN model was found to be 89.42% and 81.23%, while the prediction accuracy computed at 80.50% and 77.36%, respectively for the same phases.

However, the low results of the DNN model underline the importance of conducting further research in order to explore if other parameters, such as the amount and the quality of data, can improve its prediction performance. This will assist further the deployment of learning/assisting tools that will encapsulate machine learning, which will help further the development of more intelligent tutoring systems.

## References

1. Alshareet, O., Itradat, A., Doush, I.A., Quttoum, A.: Incorporation of ISO 25010 with machine learning to develop a novel quality in use prediction system (QIUPS). *Int. J. Syst. Assur. Eng. Manag.* **9**(2), 344–353 (2018). <https://doi.org/10.1007/s13198-017-0649-x>
2. Aouine, A., Mahdaoui, L., Moccozet, L.: A workflow-based solution to support the assessment of collaborative activities in e-learning. *Int. J. Inf. Learn. Technol.* **36**, 124–156 (2019)
3. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system, pp. 785–794 (2016)
4. Chin, K.Y., Ko-Fong, L., Chen, Y.L.: Effects of a ubiquitous guide-learning system on cultural heritage course students' performance and motivation. *IEEE Trans. Learn. Technol.* **13**, 52–62 (2019)
5. Devan, P., Khare, N.: An efficient XGBoost–DNN-based classification model for network intrusion detection system. *Neural Comput. Appl.* **32**(16), 12499–12514 (2020). <https://doi.org/10.1007/s00521-020-04708-x>
6. Dunnette, M.D., Fleishman, E.A.: *Human Performance and Productivity: Volumes 1, 2, and 3.* Psychology Press, Taylor and Francis (2014)
7. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. In: *Proceedings of International Conference on Learning Representations* (2015)
8. Mengoni, P., Milani, A., Li, Y.: Clustering students interactions in elearning systems for group elicitation. In: Gervasi, O., et al. (eds.) *ICCSA 2018. LNCS*, vol. 10962, pp. 398–413. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-95168-3\\_27](https://doi.org/10.1007/978-3-319-95168-3_27)
9. O'Donnell, A.M., Hmelo-Silver, C.E., Erkens, G.: *Collaborative Learning, Reasoning, and Technology.* Routledge, Milton Park (2013)
10. Petkovic, D., et al.: SETAP: software engineering teamwork assessment and prediction using machine learning. In: *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, pp. 1–8. IEEE (2014)
11. Petkovic, D., et al.: Using the random forest classifier to assess and predict student learning of software engineering teamwork. In: *2016 IEEE Frontiers in Education Conference (FIE)*, pp. 1–7. IEEE (2016). <https://archive.ics.uci.edu/ml/datasets/Data+for+Software+Engineering+Teamwork+Assessment+in+Education+Setting>
12. Troussas, C., Giannakas, F., Sgouropoulou, C., Voyiatzis, I.: Collaborative activities recommendation based on students' collaborative learning styles using ANN and WSM. *Interact. Learning Environ.* 1–14. Taylor and Francis
13. Troussas, C., Krouska, A., Giannakas, F., Sgouropoulou, C., Voyiatzis, I.: Automated reasoning of learners' cognitive states using classification analysis, pp. 103–106 (2020)
14. Troussas, C., Krouska, A., Giannakas, F., Sgouropoulou, C., Voyiatzis, I.: Redesigning teaching strategies through an information filtering system, pp. 111–114 (2020)
15. Wang, C., Fang, T., Gu, Y.: Learning performance and behavioral patterns of online collaborative learning: impact of cognitive load and affordances of different multimedia. *Comput. Educ.* **143**, 103683 (2020)
16. Zacharis, N.Z.: Predicting student academic performance in blended learning using artificial neural networks. *Int. J. Artif. Intell. Appl.* **7**(5), 17–29 (2016)