Prediction of Instructor Performance using Machine and Deep Learning Techniques

Basem S. Abunasser[1], Mohammed Rasheed J. AL-Hiealy[2]
Alaa M. Barhoom[3], Abdelbaset R. Almasri[4], Samy S. Abu-Naser[5]
University Malaysia of Computer Science & Engineering (UNIMY), Cyberjaya, Malaysia[1, 2, 3, 4]
Faculty of Engineering and Information Technology, Al-Azhar University, Gaza, Palestine[5]

***Abstract*—The quality of instructors' performance mainly influences the quality of educational services in higher educational institutions. One of the major challenges of higher educational institutions is the accumulated amount of data and how it can be utilized to boost the academic programs quality. The recent advancements in Artificial Intelligence techniques, including machine and deep learning models, have led to the expansion in practical prediction for various fields. In this paper, a dataset was collected from UCI Repository, University of California, for the prediction of instructor performance. In order to find how effective the instructor in the higher education systems is, a group of machine and deep learning algorithms were applied to predict instructor performance in higher education systems. The best machine-learning algorithm was Extra Trees Regressor with Accuracy (98.78%), Precision (98.78%), Recall (98.78%), F1-score (98.78%); however, the proposed deep learning algorithm achieved Accuracy (98.89%), Precision (98.91%), Recall (98.94%), and F1-score (98.92%).**

Keywords—Education; deep learning; machine learning; prediction; instructor performance

I. INTRODUCTION

Machine learning is a subclass of artificial intelligence (AI) [1, 2]. It is concerned with the teaching computers how to learn from different types of data and to enhance with experience without programming explicitly to do that [3]. In machine learning, models are being trained to look for correlations and patterns in huge datasets and be able to make predictions and decisions according to its analysis [4]. Applications of machine learning are improved with usage and become more precise when there are more data at hand. Machine learning applications are everywhere –in our offices, our supermarkets, our social media, and our hospitals [5].

Deep learning are subclass of Machine learning. Deep learning are networks with a great number of layers [6]. These layers can process broad quantities of data and find the weight of associated link in a network; such as, in an image of bird species recognition, part of the layers in the network can detect singular features in the bird's face, such as beak or eyes, whereas another layer could tell if the features in some way designate bird face [7]. Deep learning emulates how human brain operate. A few examples of deep learning are self-driven cars, medical diagnoses through sounds, classification of fish species, detection of different diseases from the eyes of the person [8]. If a network has more layers, these layers can perform complex tasks. Deep learning algorithms need high computer power to be able to produce results [10, 13].

Higher education systems claim new methodologies that increase the achievement, quality and efficiency [14, 15]. Typically Machine and Deep learning algorithms are applied in higher education for examining the effect of educational approaches on students, and in what way students comprehend the course material [11, 22]. The academic performance of students usually is based on some features like the Cumulative Grade Point Average (CGPA), economic situation, demographic data, family background and the models for prediction. Therefore the majority of the research in this field depends on the attributes that are related to the students [12, 21].

This paper is an attempt to analyze the data associated with the evaluation of the student for instructors to enhance the quality of higher educational systems and specify the factors that impact the performance of the students. Student performance prediction is largely associated with the quality of teaching.

In this study, various data prediction techniques are carried out on the student evaluation dataset for the prediction of student accomplishment, inspect instructors' performance, and discover the best technique for classification in line with these measures: Accuracy, Precision, Recall, F1-score and time performance [16].

II. LITERATURE REVIEW

It is required to measure the instructors' performance to boost the effectiveness of teaching and enhance the knowledge of students in the field of higher education. It is done by using feedback gathered from the students.

Agaoglu [1] measured the performances of instructor dependent on view of student using questionnaire of a course evaluation using four techniques: Support Vector Machines, Decision Tree algorithms (C5.0, CART), Discriminant Analysis and Artificial Neural Networks (ANN-2QH, ANN-3QH, ANNMH). The performance measurements applied were recall, precision, accuracy, and specificity. Also, feature importance was done to eliminate the irrelevant features. At last this work indicated the expressiveness and adequacy of models of data mining in higher education. Dataset was collected from Marmara University, Istanbul, Turkey. The dataset contains 2850 records, 25 features, and one class name. Among various strategies C5.0 achieved high accuracy of 92.3%.

The research in Ahmadi and Ahmad [2], inspected the attributes of instructors training performance utilizing two techniques: stepwise regression and decision trees. The dataset in that study was collected from learners of MIS department during the time of 2004-2009. Factor investigation was applied to reveal free factors influencing the overall performance of instructors. Stepwise regression model was created using SPSS and decision trees were built using Answer Tree.

Ahmed et al. [3] inspected the parts which are mainly influencing the success of learners for predicting the performance of instructors to upsurge the quality of the educational system using several techniques like Multilayer Perception (MLP), J48 Decision Tree, Sequential Minimal Optimization (SMO) and Naïve Bayes (NB). The Dataset was collected from (UCI) Repository. There were 32 features collected from Q1 to Q28 asked with responses from 1 to 5. Features were assessed utilizing R, eight features were chosen with high impact. Algorithms were applied with all features and with strongly affected features only. J48 achieved 84.8% with all the features and SMO achieved 85.8% for chosen features.

Ola and Sellapan [16] examined the feedback from students about the instructors to form Instructor Evaluation framework using WEKA tool. Data collected from 830 undergraduate studies around 104 records with five attributes. Decision tree algorithm applied and the outcomes were utilized by the educationalist to distinguish whether specific instructor is proceeded to the following semester or not. The researchers used an intelligent approach for the assessment of instructors' performance in higher institutions and proposed an optimal machine learning algorithm to design a system framework. Formative and Summative assessment methods applied to assess the instructor's performance to increase the quality.

Kumar and Saurabh [17] created a framework to predict the performance of instructors utilizing their assessment, checking the classes and performance assessment of instructors. The strategies utilized in that system was Naive Bayes, ID3, LAD tree and CART in WEKA tool. Three years of data gathered from post graduate students with 14 attributes. The precision created by ID3 was 65.14%, 72.32% by CART, 75.00% by LAD Tree and the most accuracy 80.35% produced by NB classifier.

In the study of Vijayalakshmi et al. [19] some of the machine learning algorithms were applied like "Naïve Bayes, K-Nearest Neighbor, Random Forest, Support Vector Machine, and Decision Tree". The implementation language is R programming language for data mining apps. Various implementation measures were applied to assess the system such as accuracy, precision, recall, specificity, sensitivity. The highest accuracy attained was by SVM. It was better than other models using the dataset at hand.

Yahya et al. [20] examined the practicality of applying Data Mining techniques to distinguish the practicality of instructors. Data was gathered, nine methods were applied such K-Nearest Neighbors (KNN), Naive Bayes (NB), Support Vector Machine (SVM), RA, J48, JPip, AdaBoost, BN, Random Forest among these Random Forest (RF) and SVM showed prominent implementation.

A. *Summary of the Previous Studies*

In Table I, a summarization of the above discussed previous studies in terms of Machine Learning methods, best method, tools used, accuracy of each method used, and the size of the dataset and number of attributes.

Therefore, in the current, the same dataset as in [3] which was collected from UCI Machine Learning Repository from the University of California for the prediction of instructor performance [18] was used.

TABLE I. SUMMARY OF PREVIOUS STUDIES

Reference	Techniques used	Tool Used	Results		Dataset
Agaoglu [1]	ANN, DA,C5.0,CART, SVM	IBM SPSS Modeler-Quick and Multiple Classifiers	C5.0	92.3	2850 data, 26 Attributes
			CART	89.9	
			SVM	91.3	
			ANN-Q2H	91.2	
			ANN Q3H	90.8	
			ANN-M	90.5	
			DA	90.5	
Ahmadi and Ahmad [2],	Decision tree with J48	WEKA	J48	82.60	104 records 5 features
Ahmed *et al.* [3]	MLP, DT, NB, ETR	WEKA	DT	84.8	5820 data, 33 Attributes
			NB	83.3	
			ETR	84.5	
			MLP	82.5	
Asanbe *et al.* [8]	MLP, ID3, C4.5	WEKA	ID3	71.00	2010-2015 data, 350 records, 12 Attributes
			C4.5	83.5	
			MLP	82.5	

Ola and Sellapan [16]	Decision tree with J48	WEKA	J48	71.20	830 undergraduate studies around 104 records with 5 attributes
Kumar and Saurabh [17]	NB, ID3, CART, LAD	WEKA	NB	80.35	3 Years data 14 Attributes
			ID3	65.17	
			CART	72.32	
			LAD	75.00	
Vijayalakshmi *et al.* [19]	NB, KNN, RF, SVM, C5.0	R	NB	87.7	2220 data, 21 Attributes
			KNN	91.7	
			C5.0	94.2	
			RF	98.09	
			SVM	99.25	
Yahya *et al.* [20]	KNN , NB, SVM, RA, J48, JPip, AdaBoost, BN, RF	WEKA	KNN	57.90	7348 questions 6 attributes
			NB	57.40	
			SVM	70.80	
			RA	64.90	
			J48	72.60	
			JPip	25.00	
			AdaBoost	64.70	
			BN	20.10	
			RF	55.70	

III. Methodology

This section will present the methodology of our study which includes Data Collection, Data Preprocessing, Explanation of the proposed models used for analysis and prediction.

A. Dataset

Dataset used in this study was collected from UCI Repository for the prediction of instructor performance [18]. It has 5820 with 33 features.

The features in the dataset: "instr code, class level, number of repeating the course, attendance, difficulty level, and 28 question (Q1 to Q28). Q1-Q28 are all Likert-type, meaning that the values are taken from 1-5, where 1,2,3,4,5 represents 'Poor', 'Fair', 'Good', 'Very Good', and 'Excellent' respectively for Q1 to Q28. Furthermore, there is one class variable (Performance). Performance was calculated by taking the average of the 28 Question values. The calculated values of these questions are in the form {1, 2, 3, 4, 5}, where 1,2,3,4,5 represents the 'Poor', 'Fair', 'Good' 'Very Good', and 'Excellent' respectively." [18]

B. Data Analysis

The class (performance) variable predicts the performance of the instructor. The possible values for performance are 'Poor' (888 Observations),".Fair" (996 Observations), Good" (2073 Observations), "Very Good" (1275 Observations), 'Excellent' (588 Observations). The class (performance) variable distribution is shown in Fig. 1.

C. Data Preparation

All of the 33 features and the Performance class of the dataset are already label encoded. The class (Performance) balancing was checked and found that the class is not balanced as in Fig. 1. So, Smote function was used to balance the class (Performance). The Smote function increases the number of samples of the low counts to be the same as the higher count.

D. Dataset Splitting

The dataset was split into 3 datasets: (Training, testing and validating datasets). The ratio of splitting was (60%, 20%, and 20%).

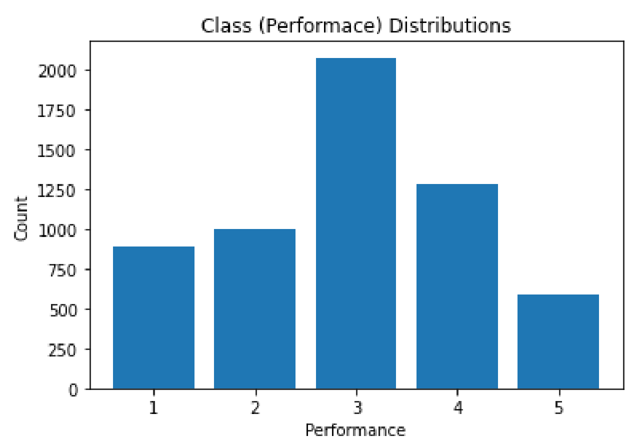

Fig. 1. Class (Performance) Distribution.

E. Description of Models used in the Study

There are many algorithms of ML that can be used in the prediction of instructor Performance level. ML algorithms were trained and tested using the current dataset with 18 various features. The algorithms that were used for prediction and analysis belong to 10 categories of Machine Learning [9] including:

- Naive_Bayes (GaussianNB).
- Neighbors (NearestCentroid, KNeighborsClassifier).
- Linear_model(LogisticRegression,LogisticRegressionCV , LinearRegression).
- SVM (SVC).
- Tree (ExtraTreeClassifier, DecisionTreeClassifier).
- XGBoost (XGBClassifier).
- Ensemble (GradientBoostingClassifier, GradientBoostingRegressor, AdaBoostRegressor, Extra TreesRegressor, BaggingClassifier, RandomForest Classifier).
- Neural_Network(MLPClassifier, MLPRegressor).
- Lightgbm(LGBMClassifier).
- Semi_supervised(LabelPropagation).

Furthermore, a deep learning model was proposed to predict instructors' performance in higher education systems. The DL proposed model consists of seven Dense layers: one input layer (33 features), five hidden layers (256,128, 64, 32, and 16 neurons), and one output layer with five classes and softmax function as can be seen in Fig. 2. The reason for using five hidden layers is the high accuracy. The structure of the DL model gave the best accuracy compared to four, three, or two hidden layers.

The steps of the methodology used in the study for predicting instructors' performance in higher education systems are summarized in Fig. 3.

```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 33)]              0

 dense (Dense)               (None, 256)               8704

 dense_1 (Dense)             (None, 128)               32896

 dense_2 (Dense)             (None, 64)                8256

 dense_3 (Dense)             (None, 32)                2080

 dense_4 (Dense)             (None, 16)                528

 dense_5 (Dense)             (None, 5)                 85

=================================================================
Total params: 52,549
Trainable params: 52,549
Non-trainable params: 0
_________________________________________________________________
```

Fig. 2. Structure of the Proposed Deep Learning Model.

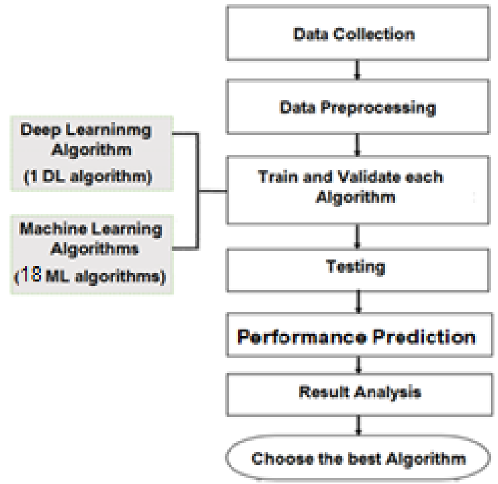

Fig. 3. Methodology for the Prediction of Instructors' Performance in Higher Education Systems.

IV. Result and Analysis

In the following sections, a discussion of the result achieved by the deep learning model and the machine learning models will be presented. The first section talks in detail about the Performance Evaluation and the second section presents the Performance Analysis of all models used in this study.

V. Performance Evaluation

Different machine and deep learning measurements can be applied on the various model used in the current study. The most popular measurements are: Accuracy, F1- score, Recall and Precision are the most important criterion used to assess a models performance. The value of the confusion matrix which is generated during the testing of the model is considered to calculate these measurements as illustrated in equation 1, 2, 3 and 4.

$$\text{Accuracy} = (TP + TN) / (TP+TN+FP+FN) \quad (1)$$

$$\text{Precision} = (TP / (TP+FP) \quad (2)$$

$$\text{Recall} = (TP / (TP+FN) \quad (3)$$

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

Where TP = True Positive, TN = True Negative

FP = False Positive, FN = False Negative

VI. Performance Analysis of Applied Models

In this study, 18 machine learning algorithms to predict instructors' performance in higher education systems were used. Furthermore, a deep learning model was proposed to predict instructors' performance in higher education systems. The aim of this study was to get a more efficient predictive model by making a comparison between the different deep and machine learning models. 60% of the dataset for training, 20% of the dataset for validating and the remaining 20% of the dataset were used for the testing.

TABLE II. PERFORMANCE OF THE MACHINE AND DEEP LEARNING ALGORITHMS

Model Type	Model Name	Accuracy	Precision	Recall	F1-score	Time in Sec
Machine Learning	Extra Trees Regressor	**98.78%**	**98.78%**	**98.78%**	**98.78%**	2.70
	Gradient Boosting Regressor	97.71%	97.71%	97.71%	97.71%	0.80
	Random Forest Classifier	97.35%	97.36%	97.35%	97.35%	0.65
	Logistic Regression CV	97.25%	97.25%	97.25%	97.25%	308.06
	LGBM Classifier	97.01%	97.01%	97.01%	97.01%	1.10
	Gradient Boosting Classifier	95.95%	95.95%	95.95%	95.95%	5.55
	MLP Classifier	95.80%	95.80%	95.80%	95.80%	11.37
	Bagging Classifier	95.75%	95.75%	95.75%	95.75%	0.25
	Logistic Regression	94.64%	94.64%	94.64%	94.64%	1.93
	Extra Tree Classifier	93.83%	93.83%	93.83%	93.83%	0.02
	Decision Tree Classifier	93.39%	93.39%	93.39%	93.39%	0.03
	Label Propagation	93.25%	93.25%	93.25%	93.25%	2.02
	Gaussian NB	92.47%	92.47%	92.47%	92.47%	0.02
	K Neighbors Classifier	91.66%	91.66%	91.66%	91.66%	0.41
	SVC	91.66%	91.66%	91.66%	91.66%	0.61
	Linear Discriminant Analysis	91.27%	91.27%	91.27%	91.27%	0.07
	Ada Boost Regressor	91.16%	91.16%	91.16%	91.16%	2.12
	Nearest Centroid	90.79%	90.79%	90.79%	90.79%	0.02
Deep Learning	Proposed Deep Learning Model	**98.89%**	**98.91%**	**98.94%**	**98.92%**	**2.00**

To evaluate the models performance, five sorts of assessment measures were engaged: “Recall, Precision, Accuracy, F1-Score and time needed for each model to run are shown in Table II. It is observed that the best machine-learning algorithm was “Extra Trees Regressor” with an Accuracy (98.78%), Precision (98.78%), Recall (98.78%), and F1-score (98.78%); however, the proposed deep learning algorithm achieved an Accuracy (98.89%), Precision (98.91%), Recall (98.94%), and F1-score (98.92%).

VII. RESULT AND DISCUSSION

All previous studies reviewed in the section of literature review except one used different datasets; thus the results of these studies cannot be compared with the result obtained in the current study.

The previous study that used the same dataset as in the current study is Ahmed et al. [3]. The following table compares their results with the current proposed model’s results.

As it can be seen in Table III, the results of the current study are much higher than the results obtained in the previous study.

TABLE III. RESULTS COMPARISON WITH PREVIOUS STUDIES

Model Name	**Previous study (Ahmed *et al.* [3])**	**Proposed Models of the Current Study**
Decision Tree Classifier (DT)	84.80	93.83
Gaussian NB	83.30	92.47
Extra Trees Regressor (ETR)	84.50	98.78
MLP Classifier	82.50	95.80

The reason for the high accuracy of the current study is the pre-processing the handling of the dataset.

VIII. CONCLUSION

In this study, 18 different Machine Learning algorithms and one deep learning algorithm for predicting instructors’ performance in higher education systems were used. The dataset was collected from UCI Repository for the prediction of instructor performance. The dataset was preprocessed, the class (performance) was balanced using smote function. Each algorithm was trained, tested and its performance was noted. Furthermore, the proposed deep learning model was trained, validated and tested using the same dataset and its performance was noted. Among all the machine learning models used, the best machine-learning algorithm was Extra Trees Regressor with an Accuracy (98.78%), Precision (98.78%), Recall (98.78%), F1-score (98.78%); however, the proposed deep learning algorithm achieved an Accuracy (98.89%), Precision (98.91%), Recall (98.94%), F1-score (98.92%). Even though, the accuracies of the best machine learning algorithm and the proposed deep learning algorithm were close; the proposed deep learning algorithm was slightly better.

These discoveries are helpful to educationalist to improve their performances.

REFERENCES

[1] Agaoglu, Mustafa. "Predicting Instructor Performance Using Data Mining Techniques in Higher Education." IEEE Access, 2016, Vol. 4, pp. 2379-2387.

[2] Ahmadi, Fateh, and ME Shiri Ahmad. "Data Mining in Teacher Evaluation System using WEKA." International Journal of Computer Applications, 2013, Vol. 63, No. 10.

[3] Ahmed, Ahmed Mohamed, Ahmet Rizaner, and Ali Hakan Ulusoy. "Using data mining to predict instructor performance." Procedia Computer Science, 2016, Vol. 102, PP. 137-142.

[4] Saleh, A., Sukaik, R., Abu-Naser, S.S. Brain tumor classification using deep learning. Proceedings - 2020 International Conference on Assistive and Rehabilitation Technologies, iCareTech 2020, 2020, pp. 131–136, 9328072.

[5] Arqawi, S., Atieh, K.A.F.T., Shobaki, M.J.A.L., Abu-Naser, S.S., Abu Abdulla, A.A.M. Integration of the dimensions of computerized health information systems and their role in improving administrative performance in Al-Shifa medical complex, Journal of Theoretical and Applied Information Technologythis link is disabled, 2020, Vol. 98, No. 6, pp. 1087–1119.

[6] Buhisi, N. I., & Abu-Naser, S. S. Dynamic programming as a tool of decision supporting. Journal of Applied Sciences Research, 2009, Vol. 5, No. 6, pp. 671-676.

[7] Naser, S. S. A. Developing visualization tool for teaching AI searching algorithms. Information Technology Journal, 2008, Vol. 7, No. 2, pp. 350-355.

[8] AlKayyali, Z. K. D., et al. Prediction of Student Adaptability Level in e-Learning using Machine and Deep Learning Techniques. International Journal of Academic and Applied Research (IJAAR), 2020, Vol. 6, No. 5, pp.84-96.

[9] Naser, S. S. A. Intelligent tutoring system for teaching database to sophomore students in Gaza and its effect on their performance. Information Technology Journal, 2006, Vol. 5, No. 5, pp. 916-922.

[10] Mady, S.A., Arqawi, S.M., Al Shobaki, M.J., Abu-Naser, S.S. Lean manufacturing dimensions and its relationship in promoting the improvement of production processes in industrial companies. International Journal on Emerging Technologies, 2020, Vol. 11, No. 3, pp. 881–896.

[11] Albatish, I.M., Abu-Naser, S.S. Modeling and controlling smart traffic light system using a rule based system. Proceedings - 2019 International Conference on Promising Electronic Technologies, ICPET 2019, 2019, pp. 55–60.

[12] Elzamly, A., Messabia, N., Doheir, M., ...Al-Aqqad, M., Alazzam, M. Assessment risks for managing software planning processes in information technology systems. International Journal of Advanced Science and Technology, 2019, Vol. 28, No. 1, pp. 327–338.

[13] Abu Ghosh, M.M., Atallah, R.R., Abu Naser, S.S. Secure mobile cloud computing for sensitive data: Teacher services for palestinian higher education institutions. International Journal of Grid and Distributed Computing, 2016, Vol. 9, No. 2, pp. 17–22.

[14] Elzamly, A., Hussin, B., Naser, S.A., ...Selamat, A., Rashed, A. A new conceptual framework modelling for cloud computing risk management in banking organizations. International Journal of Grid and Distributed Computing, 2016, Vol. 9, No. 9, pp. 137–154.

[15] Abu-Naser, S.S., El-Hissi H., Abu-Rass, M., & El-khozondar, N. An expert system for endocrine diagnosis and treatments using JESS. Journal of Artificial Intelligence, 2010, Vol. 3, No. 4, pp. 239-251.

[16] Abu Naser, S.S. Evaluating the effectiveness of the CPP-Tutor, an intelligent tutoring system for students learning to program in C++. Journal of Applied Sciences Research, 2009, Vol. 5, No. 1, pp. 109-114.

[17] V. Vijayalakshmi, K. Panimalar, S. Janarthanan. “Predicting the performance of instructors using Machine learning algorithms.” High Technology Letters, 2022, Vol. 26, No. 12, pp. 694-705.

[18] UCI Machine Learning Repository, https://archive.ics.uci.edu/.

[19] Zaqout, I., et al. "Predicting Student Performance Using Artificial Neural Network: in the Faculty of Engineering and Information Technology." International Journal of Hybrid Information Technology, 2015, Vol. 8, No. 2, pp. 221-228.

[20] Obaid, T., Eneizan, B., Naser, S.S.A., ...Abualrejal, H.M.E., Gazem, N.A. Factors Contributing to an Effective E- Government Adoption in Palestine. Lecture Notes on Data Engineering and Communications Technologies, 2022, 127, pp. 663–676.

[21] Naser, S. S. A. JEE-Tutor: An intelligent tutoring system for java expressions evaluation. Information Technology Journal, 2008, Vol. 7, No. 3, 528-532.

[22] Naser, S. S. A. Developing an intelligent tutoring system for students learning to program in C++. Information Technology Journal, 2008, Vol. 7, No. 7, pp. 1051-1060.