

# Educational Data Mining: A Predictive Model for Cisco Certification Exam using Classification Algorithms

Nenita D. Guerrero

## Abstract

This study aims to identify students who are vulnerable of not being able to pass the Cisco certification examination. The main goal is to develop a model that will determine the significant attributes that influence students' success in Cisco certification examination. The significant attributes were determined using logistic regression. The researcher conducted preliminary interviews in selected Cisco academies to determine prevailing issues. The study used sets of classification algorithms to generate models that were used for prediction. The main function of the model is to predict the probability of the examinee to pass a Cisco certification examination. The researcher used data mining tools such as WEKA and SPSS to derive the required models. Various data mining classification algorithms were used to identify the most accurate technique best suited for the given data set. The result of the experiment showed that the Logistic Regression algorithm is the most accurate algorithm to be used in the development of the predictive model.

Keywords:

data mining, classification algorithm, logistic regression, predictive model, students' performance

**Suggested Citation:** Guerrero, N.D. (2021). Educational Data Mining: A Predictive Model for Cisco Certification Exam using Classification Algorithms. *The Research Probe*, Volume 1, Issue 1, pp 1 – 17.

About the author:

Assistant Professor, Abu Dhabi Polytechnic, United Arab Emirates

## 1. Introduction

In this era, the rapid growth of networking technology has been constantly increasing worldwide to promote economic development. This increases the global demand for highly skilled network professionals. Cisco, being known as the leading vendor of networking services and equipment, has opted to put up a networking academy to provide intensive training to aspiring network professionals. The training provided develops the required knowledge and skills to implement and maintain networking solutions. It also prepares students to be ready for an equivalent certification examination. Courses in the academy are offered through blended learning that combines classroom instruction with online curricula, interactive tools, hands-on activities, and online assessments that provide immediate feedback. These courses are essential to pass a Cisco certification.

This study aims to develop a predictive model using various machine learning classification algorithms that will identify students who need remediation or review plans to further improve the chance of passing the examination. The study will assist Cisco academy instructors to identify important and relevant subjects or attributes that have a significant contribution to pass the certification exam.

The generated model focuses on the early prediction of students who will have difficulty and low chance of passing the examination, thus appropriate support and reviews can be administered by the institution involved. The predicted value will serve as their guide in the design of their teaching materials and methodology in an approach that is best suited to the abilities of the students they handle.

Specific problems that these research addresses are as follows:

1. What are the significant attributes that contribute to the prediction of examinees' success in Cisco certification?
2. What data mining classification technique is the most accurate in predicting students' academic performance in the Cisco certification exam?

The outcome of this study is not intended to be the sole source of the decision in the evaluation of students' performance; instead, it will serve as a supplementary tool in the evaluation and analysis of students' learning achievement in Cisco certification.

## **2. Literature Review**

### **2.1. Data Mining**

The ability to predict a student's performance is very vital in educational institutions. Every student's performance could be based on diverse factors such as personal, academic, social, psychological, and other environmental factors. According to Sree & Rupa (2013), this objective could be attained through the use of data mining techniques.

Friedman (2009) defined data mining as an interdisciplinary subfield of computer science, in which its main goal is to extract hidden patterns or data models present in a database using machine learning algorithms. The main goal of data mining is to extract information from a data set and transform it into a meaningful structure for further use.

### **2.2. Data Mining in an Educational Context**

Data Mining is used in the educational field to enhance our understanding of the learning process to focus on identifying, extracting, and evaluating variables related to the learning process of students. Sonali et al. (2012) determined that data mining could be used to improve the education system and improve the service and overall efficiency by optimizing the resources available.

On the other hand, Kumar (2011) discussed that educational data mining is used to study the data available in the educational field and discover the hidden knowledge from it. He mentioned that classification techniques can be applied on the data for predicting student's performance.

Alaa el-Halees (2009) determines that educational data mining is one of the key areas in data mining that is gaining popularity because of its potential to extract educational patterns suitable and necessary to student, faculty, and administration behavior and performances. Data Mining can be used in the educational field to enhance our understanding of the learning process to focus on identifying, extracting, and evaluating variables related to the learning process of students.

Educational data mining is an interesting research area that extracts useful, previously unknown patterns from the educational database for better understanding, improved educational performance, and assessment of the student learning process (Surjeet & Saurabh, 2012).

In the last few years, researchers have already begun to apply various data mining methods to help teachers improve e-learning systems (Romero and Ventura, 2006).

Kumar (2011) discussed that educational data mining is used to study the data available in the educational field and bring out the hidden knowledge from it. Classification methods like decision trees, rule mining, Bayesian network, and regression can be applied to the educational data for predicting the students' behavior, performance in the examination, etc. This prediction will help the tutors to identify the weak students and help them score better marks.

According to Romero and Ventura (2010), educational data mining (EDM) has emerged as a new field of research, capable of exploiting the abundant data generated by various systems for use in decision making. The enthusiastic adoption of data mining tools by higher education has the potential to improve some aspects of the quality of education, while it lays the foundation for a more effective understanding of the learning process.

### ***2.3. Data Mining applications in Education***

Bhardwaj and Pal (2011) selected 300 students from five different degrees. The researcher utilized Bayesian classification method on 17 attributes. The study reveals that factors like students' grades in senior secondary examination, location or residency, medium of teachers' instructional competence, students' habits, family annual income, and students' family were highly significant and related in predicting student academic performances.

Pandey and Pal (2011) conducted a study on student academic performance by selecting 60 students from different colleges of Dr. R. M. L. Awadh University, Faizabad, India. The researcher used association rule or apriori algorithm to identify and find the interestingness of students in opting class teaching. The algorithm enables to identification best association rules found on the class domain.

Mohammed M. Abu Tair and Alaa M. El-Halees (2012) adopted educational data mining techniques by developing data models in which data models are knowledge discovered from the educational domain. The data models are being used to understand and improve students' academic performances and overcome the problem of low grades. The researcher used data within fifteen years period [1993-2007]. The researcher used the pre-processing technique before the application of association and classification algorithms to determine predictive models including equations and rule sets.

Tarun et. al. (2014) presented the integration of data mining and decision support systems in an educational context, resulting in a predictive decision support system for licensure examination performance. The researchers integrated a classification data model derived from multiple regression and PART classification techniques. The researcher presented a model for integration of decision support and data mining by having a framework called PDSS-LEP. The model was found beneficial as it provides a good platform for the generation of the MR model that can be adapted by other institutions because of its model selection procedures and user-oriented interface. It is, however, suggested that data integration should be enhanced by considering multiple sources of data.

#### **2.4. Application of Classification Algorithms**

Pandey (2011) defined classification or supervised learning as the most applied on predicting data sets. The goal of the classification algorithm is to create a data model that can predict and classify unclassified data records. The process includes two kinds of steps: learning and classification. In the learning step, the training data set is analyzed by the classification algorithm. The training data set is used to approximate the precision of classification rules. The pre-classified data records are used by the classifier training algorithm to conclude the required parameters for proper identification/discrimination.

In another study conducted by Gorikhan (2016), prediction models were developed using classification techniques such as decision tree, neural network, logistic regression, support vector, and neural networks. The outcome of these models is to predict the number of students who were likely to pass or fail. The results were given to teachers and steps were taken to improve the academic performance of the weak/failing students. After analysis and comparison, it was found that the model generated decision tree analysis and logistic regression recorded the highest accuracy rates incorrectly classified prediction.

Brijesh and Saurabh (2011) concluded that variables such as semester marks and attendance can be used as attributes in the classification techniques for predicting end semester results.

García-Saiz and Zorrilla (2011) focused on reviewing the strategy by looking at the performance of the students at Junior Secondary Certificate examinations in the Ondo State, Nigeria. In one of the experiments done for evaluating the performance of various

classification techniques for distance education students' education dataset, it has been identified that I Bayes performed adequately with an accuracy of 80.97%.

Romero et al.( 2008) cited that classification algorithm is one of the most widely used data mining techniques used by different researchers for data analysis and investigation if there are hidden patterns stored in a database. The classification algorithm is considered a supervised learning approach where the class labels are defined. Classification uses training record sets with labeled attributes that are used for designing data models in order to predict unknown records (Baradwaj & Pal, 2012).

Nguyen and Peter (2007) conducted a study of two different groups of students, including both undergraduate and post-graduate levels. The main objective of the study is to predict the performances of students and to compare the efficiency of two classifiers including decision tree and Naïve Bayes algorithms. The processing and modeling of data models were processed using the WEKA tool. In this research, the performance of the Decision tree was 3-12% more accurate than Bayesian networks. This was useful for identifying the weak students for further guidance and for selecting good students for the scholarship.

R. R. Kabra and Bichkar (2011) created a study focusing on 346 engineering students studying in their first year. The goal of the study is to develop a classification model based on their past academic performances. A two-class prediction and three-class prediction have been compared under the study. The results of two-class predictions were better than a three-class prediction, which helped identify the students who would likely fail.

Pittman (2008) performed a study to explore the effectiveness of data mining methods in identifying students who are at risk of leaving a particular institution. The study also aimed to compare data mining methods and techniques for students' classification based on their module usage data and the final marks in their respective programs or courses. The study identifies that the most appropriate algorithm was decision trees, for being accurate and comprehensible for instructors.

Kabakchieva (2011) also developed models for predicting student performance, based on their personal, pre-university, and university performance characteristics.

Gorikhan (2016) emphasized techniques in data mining in the development of data models that will predict the academic performance of students using attributes of their grades in math and science from previous examinations. The prediction models were developed using

classification techniques such as decision trees, neural networks, logistic regression, support vector, and neural networks. The outcome of these models is to predict the number of students who were likely to pass or fail. The results were given to teachers and steps were taken to improve the academic performance of the weak/failing students. After analysis and comparison, it was found that the model generated decision tree analysis and logistic regression recorded the highest accuracy rates incorrectly classified prediction.

The study conducted by Kumar (2014) aimed to predict the third-semester grade of students enrolled in MCA. The researcher used decision tree algorithms which include J48 and Random Tree, to build rule sets or data models. The study used potential 250 students considering 20 attributes. The attributes consist of a combination of academic integration, social integration, and emotional skills. The researcher utilized an open-source data mining tool, Waikatao Environment for Knowledge Analysis, or popularly known as WEKA. The WEKA provides a build algorithm to generate a data model based on the train sets. The performance of the algorithm was evaluated based on the recall and precision and true positive rate. Precision is defined as number 250 with correct positive prediction over a total number of positive predictions. However, recall is defined as the number of correct positive predictions over a total number of positive cases. A high precision indicates that the algorithm returns more relevant results than irrelevant and high recall means that most of the results returned by the algorithm are relevant. The paper had proven that the second semester was very relevant to the academic success of students entering the third semester. On the same note, programming subjects of the second semester formed the foundation of programming subjects of the third semester. In conclusion, good academic performance in previous semesters is an indicator of the future performance of student's academic achievements.

In the study entitled “Mining Educational Data to Analyze Students’ Performance”, one of the various ways to attain quality education in HEIs is by discovering knowledge from a data set to be used for prediction of the enrollment of students in a particular course, separation of the student from traditional teaching environment, the discovery of unfair means used in online examinations, as well as detection of anomalies in the result sheets of the students, prediction about students’ performance and so on. The knowledge needed is hidden in the educational data set and it can be extracted using data mining techniques. In this study, a classification algorithm was used to evaluate a student’s performance and among the different

approaches that are used for data classification, the decision method is used here. Knowledge was extracted and used to describe the performance of the student by the end of the semester. It aids in the early identification of dropouts and students who require special attention. It also enables the teacher to provide the necessary support required by students. (Baradwaj & Pal, 2011).

The amount of data stored in the educational database is growing rapidly. The stored data in the database contains hidden knowledge about students' performance and behavior. The ability to predict student's performance in the educational context is very vital. Student's academic performance is affected by psychological and environmental factors. This can be predicted by an appropriate educational data mining technique. (Kumar, 2014)

Many factors influence the academic performance of the students. The factors that describe student performance can be used for predicting students' performance by using a number of well-known data mining classification algorithms, such as ID3, REPTree, Simplecart, J48, NB Tree, BFTree, Decision Table, MLP, and Bayesnet. The study model mainly focused on analyzing the prediction of the academic performance of the students by all the classification algorithms. The algorithms were analyzed based on the precision of predicting the result. (Ruby & David, 2014)

### **2.5. *Synthesis of the Study***

In this study, the researcher included all potential attributes. The researcher used regression analysis as a pre-processor for predictive data mining to determine the significant attributes that contribute to the success of examinees in the Cisco certification exam.

The study is focused on labeled class utilizing various classification algorithms in processing the data. The main goal of the study is to extract hidden patterns or models that can be used to predict the success of an examiner in the Cisco networking examination.

The process of the study is divided into two categories: training and testing data. A data model will be built in the training set by lists of different classifiers under supervised learning. To empirically test which algorithm will be used, all necessary classifiers were processed.

To evaluate the results of the classifier confusion matrix or accuracy computation will be used to measure the effectiveness of the algorithm. A classification error rate was calculated for the model and stored as an independent test error rate for the first model. The misclassification table was used to evaluate the prediction of a classifier.



### 3. Methodology

The researcher conducted preliminary interviews in selected institutions that offer CNAP to identify existing issues. The study used sets of classification algorithms to generate models that were used for prediction. These models were used to determine the success rate of exam takers. The researcher used data mining tools such as WEKA and SPSS to derive the required models. The confusion matrix /confusion table was used to determine the accuracy of the model.

**Table 1**

**Confusion Matrix**

		Prediction	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

The *true positives* (TP) and *true negatives* (TN) are correct classifications. A *false positive* (FP) occurs when the outcome is incorrectly predicted as *yes* (or positive) when it is *no* (negative). A *false negative* (FN) occurs when the outcome is incorrectly predicted as negative when it is positive.

To calculate the accuracy result of the model, the following equation was used:

**Equation 1: Accuracy**

$$\frac{TP + TN}{(TP + TN + FP + FN)}$$

The accuracy results determine the right classification divided by the total number of data instances.

### 4. Results and Discussion

Table 2 indicates the coefficient and significant values of the attributes.

This research aims to generate a predictive model suitable for predicting Cisco Certification results based on a given set of parameters. Sixty percent of training data and forty

percent of test data were used in building the model. If the accuracy of the model is acceptable, the model can be used to classify future data instances for which the class label is not known. Since the target variable is a dichotomous variable consisting of binary values to determine the significant attributes, logistic regression was used. The data used normative transmutation for easy manipulation. The data consist of Cisco certification exam result as target variable, Cisco grades including Cisco final and practical examination, demographic profile, and other academic data. To determine the strength of the variables, the potential attributes were processed using the logistic regression technique.

**Table 2**  
*Values in the Equation Using Logistic Regression Analysis*

<b>Attributes</b>	<b>B</b>	<b>S.E.</b>	<b>Wald</b>	<b>df</b>	<b>P value</b>	<b>Exp(B)</b>
Gender	0.416	0.311	1.781	1	0.182	1.516
Scholarship	0.412	0.333	1.527	1	0.217	1.509
Core_Programming	-0.251	0.406	0.381	1	0.537	0.778
IT_Professional	0.959	0.375	6.533	1	0.011	0.383
IT_Elective	1.109	0.326	11.577	1	0.001	3.031
Social_Science	-0.363	0.285	1.623	1	0.203	0.696
Final_Examination_Cisco1	0.135	0.440	0.094	1	0.759	1.144
Final_Practical_Cisco1	0.650	0.452	2.068	1	0.150	0.522
Cisco1	1.676	0.294	32.455	1	0	5.344
Final_Examination_Cisco2	0.110	0.277	0.157	1	0.692	0.896
Final_Practical_Cisco2	-0.054	0.327	0.027	1	0.869	0.948
Cisco2	0.075	0.467	0.026	1	0.872	1.078
Final_Examination_Cisco3	0.577	0.402	2.062	1	0.151	1.780
Final_Practical_Cisco3	0.696	0.362	3.701	1	0.054	2.006
Cisco3	0.302	0.207	0.967	1	0.325	1.352
Final_Examination_Cisco4	-0.443	0.512	0.749	1	0.387	0.642
Final_Practical_Cisco4	0.137	0.370	0.137	1	0.712	1.147
Cisco4	0.973	0.303	10.297	1	0.001	2.646
Constant	-7.964	0.942	71.443	1	0	0

To determine the statistical significance of an attribute the  $p$ -value was used. The attribute is statistically significant when a  $p$ -value is less than the significance level. The  $p$ -value is the probability of observing an effect given that the null hypothesis is true whereas the significance or alpha ( $\alpha$ ) level is the probability of rejecting the null hypothesis given that it is true. In practice significance level is chosen before data collection and is usually set to 0.05.

**Table 3*****Significant Attributes in Predicting Cisco Certification using Regression Analysis***

Attributes	B	S.E.	Wald	df	P value	Exp(B)	95% C.I. for Exp(B)	
							Lower	Upper
IT_Professional	-.959	.375	6.533	1	.011	.383	.184	.800
IT_Elective	1.109	.326	11.577	1	.001	3.031	1.600	5.742
Cisco1	1.676	.294	32.455	1	.000	5.344	3.002	9.512
Final_Practical_Cisco3	.696	.362	3.701	1	.050	2.006	.987	4.077
Cisco4	.973	.303	10.297	1	.001	2.646	1.460	4.793

Legend: \* $p \leq 0.05$

Table 3 indicates the significant attributes in predicting Cisco Certification Examination.

The dependent variable in the analysis is Cisco certification status coded so that 1 = not pass and 2 = pass. The model was generated using SPSS. Results of the significant attributes are processed using binary logistic regression is summarized in Table 3. Analysis of the data reveals that five variables significantly predict Cisco certification status. Information Technology elective subjects, Cisco 1, final practical Cisco 3, and Cisco 4 have a positive  $\beta$  coefficient, indicating that the higher the scores of the students in the lists of attributes, the higher the likelihood that they will pass the Cisco examination. The coefficient reveals that IT Elective has a value of 1.109 coefficient where  $p$ -value  $\leq 0.05$ , IT Professional -0.959 where  $p$ -value  $\leq 0.05$  Cisco 1 has a coefficient value of 1.679 where  $p$ -value  $\leq 0.05$ , final practical Cisco 3 has a coefficient value of 0.696 where  $p$ -value  $< 0.05$  and cisco 4 coefficient value 0.937 where  $p$ -value  $\leq 0.05$ . The positive  $\beta$  coefficients of the academic subjects indicate that the

higher the grades of the students on such subjects, the higher the odds of passing the certification.

The researcher aims to determine the significant attributes that influence students' success in Cisco certification using logistic regression. The logistic Regression Model uses the Logit model. It provides an association between the independent variables and the logarithm of the odds of a categorical response variable. The target variable is a binary variable consisting of yes and no, the binary logistic regression model was used. Logistic regression estimated the chances of an examinee in passing the Cisco certification exam. The logistic function can take input with any value from negative to positive infinity, whereas the output always takes values between zero and one and hence is interpretable as a probability. The logistic function can be written as:

**Equation 2: Logistic Function**

$$F(x) = \frac{1}{1 + e^{- (\beta_0 + \beta_1 x)}}$$

Where F(x) would be interpreted as Probability of the examinee Prob (examinee) of the dependent variable equates to the probability of examinee to pass Cisco certification examination. The model that was generated from logistic regression is shown in Table 6.

The values in the equation found in Table 6 of the logistic regression values can be written in equation form.

The logistic function can take input with any value from negative to positive infinity, whereas the output always takes values between zero and one and hence is interpretable as a probability.

The logistic function can be written as:

$$\text{logistic}(x) = \frac{1}{1 + e^{-x}}$$

**Equation 3: Equation Derived from the logistic function**

$$\text{Prob (passing)} = \frac{1}{1 + e^{(-(-7.964-(0.959*X1)+(1.109*X2)+(1.676*X3)+(0.696 *X4)+(0.973*X5))}}$$

Where: X<sub>1</sub> = Information Technology Professional subjects; X<sub>2</sub> = Information Technology Electives;

$X_3 = \text{Cisco 1}$ ;  $X_4 = \text{Final Practical Cisco 3}$ ;  $X_5 = \text{Cisco 4}$

The main objective of a predictive model is to maximize the correctly classified instances. The confusion matrix evaluates the forecasting precision of a predictive model. For binary classification scenarios, the misclassification rate gives the overall model performance with respect to the exact number of categorizations in the training data.

The *true positives* (TP) and *true negatives* (TN) are correct classifications. A *false positive* (FP) occurs when the outcome is incorrectly predicted as *yes* (or positive) when it is *no* (negative). A *false negative* (FN) occurs when the outcome is incorrectly predicted as negative when it is positive. The equation below was used to calculate the result accuracy of the model:

$$\frac{\text{TP} + \text{TN}}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}$$

The accuracy results determine the right classification divided by the total number of data instances.

**Table 4**

***Accuracy Results of Different Methods of Logistic Regression***

Method	Accuracy Rate	Error Rate	Precision	Recall
Forward Condition	87.90%	12.10%	84.70%	85.09%
Forward Wald	87.90%	12.10%	84.70%	85.09%
Forward Likelihood Ration	87.90%	12.10%	84.70%	85.09%

The data shown in Table 4 indicates that all methods under the logistic regression algorithm resulted in the same accuracy performance with 87.90% correctly classified and 12.10% error classification rate in predicting Cisco certification. All methods under logistic regression recorded similar precision and recalled values which indicate that 84.70% of the retrieved instances are relevant and 85.09% relevant data instances were retrieved.

**Table 5**

*Accuracy Results of Different Methods of Decision Tree Algorithm*

<b>Method</b>	<b>Accuracy Rate</b>	<b>Error Rate</b>	<b>Precision</b>	<b>Recall</b>
CHAID	85.70%	14.30%	74.35%	85.29%
QUEST	82.80%	17.20%	76.13%	81.70%
CRT	81.90%	18.10%	79.78%	78.12%
Exhaustive Chaid	82.00%	18.00%	73.68%	78.87%
RepTree	86.06%	13.94%	86.50%	86.10%
J48	86.54%	13.46%	86.90%	86.50%
Random Forest	87.02%	12.98%	87.10%	87.00%
Random Tree	87.02%	12.98%	87.00%	87.00%

Upon testing the different methods of decision tree algorithm, Table 5 shows that Random Forest and Random Tree obtained the highest accuracy results in predicting Cisco certifications with an accuracy rate of 87.02% and an error classification rate of 12.98%. The random forest also recorded the highest positive predictive value or relevant data retrieval of 87.10% and a sensitivity value of 87.00%.

**Table 6**

*Summary of Accuracy Results of Selected Classification Algorithms*

<b>Method</b>	<b>Accuracy Rate</b>	<b>Error Rate</b>	<b>Precision</b>	<b>Recall</b>
Decision Tree	87.02%	12.98%	87.10%	87.00%
Logistic Regression	87.90%	12.10%	84.70%	85.09%

Table 6 shows the highest accuracy rates generated upon testing the data sets with the different methods under selected data mining techniques such as logistic regression and decision tree. It is evident that logistic regression generated the highest accuracy result of 87.90% with an error rate of 12.10%, positive predictive value (precision) of 84.70%, and sensitivity (recall) of 85.09%.

Based on the above result, the researcher developed a predictive model from the values in the equation processed using Logistic Regression algorithm.

## 5. Conclusion

The study aimed to explore machine learning algorithms, mainly classification algorithms in the prediction of students' performance in Cisco certification. The main goal of this research was to develop a predictive model that could identify students who are vulnerable of not being able to pass the Cisco certification examination. Since the target variable in this study is dichotomous with only two possible values (pass or fail), logistic regression was applied to determine the significant attributes that contribute to the examinees' success in Cisco certification exam. A predictive model has been developed through the derivation of an equation based on the logistic function.

## References

- Alaa el-Halees. (2009), "Mining students data to analyze e-Learning behavior: A Case Study", 2009.
- Baradwaj, B. K., & Pal, S. (2012). "Data Mining: A prediction for performance improvement using classification", *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 9, No. 4, pp. 136-140, 2011
- Baradwaj, B.K. and Pal, S. (2011). Mining Educational Data to Analyze Students' Performance. (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 6
- Brijesh, K.B. & Saurabh, P. (2011). Mining educational data to analyze students' performance. *Int. J. Adv. Comput. Sci. Appl.*, 2(6): 63-69
- Cisco Systems (2012). Retrieved May 1, 2015, from <https://www.netacad.com/about-networking-academy/>
- Friedman, Jerome (2009). "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". Retrieved 2012-08-07
- García-Saiz, D. and M. Zorrilla, 2011. Comparing classification methods for predicting distance students' performance. *Proceeding of the 2<sup>nd</sup> Workshop on Applications of Pattern Analysis Workshop and Conference*, pp: 26-32
- Gorikhan (2016). A study on Implementation of classification techniques to predict students' results for Institutional Analysis

- Kabakchieva, D. Predicting Student Performance by Using Data Mining Methods for Classification. *Cybernetics and Information Technologies*, 13(1):61–72, 2013
- Kabra R. R., Bichkar R.S. (2011). Performance Prediction of Engineering Students using Decision Trees, *International Journal of Computer Applications* (0975- 8887), Vol-36-No.11
- Kumar, A. and Vijayalakshmi M. (2011). EFFICIENCY OF DECISION TREES IN PREDICTING STUDENT’S ACADEMIC PERFORMANCES.
- Kumar A.D. et al (2014), A Survey on Predicting Student Performance. (*IJCSIT*) *International Journal of Computer Science and Information Technologies*, Vol. 5 (5), 6147-6149
- Kumar, A. and Priya, K. (2013). Improving the Student’s Performance Using Educational Data Mining. *Int. J. Advanced Networking Applications*. Volume: 04 Issue: 04, Pages 1680-1685 (2013), ISSN : 0975-0290
- Mohammed M. Abu Tair, Alaa M. El-Halees, Mining Educational Data to Improve Students’ Performance: A Case Study, 2012
- Nguyen N., Paul J., and Peter H. (2007). A Comparative Analysis of Techniques for Predicting Academic Performance. In *Proceedings of the 37<sup>th</sup> ASEE/IEEE Frontiers in Education Conference*. Pp. 7-12
- Pandey, U. K. and Pal, S. (2011) “Data Mining: A prediction of performer or underperformer using classification”, (*IJCSIT*) *International Journal of Computer Science and Information Technology*, Vol. 2(2), pp.686-690, ISSN:0975-9646
- Pandey, U. K. and Pal, S. “A Data mining view on class room teaching language”, (*IJCSI*) *International Journal of Computer Science Issue*, Vol. 8, Issue 2, pp. 277-282, ISSN: 1694-0814, 2011
- Pandey, U. K., & Pal, S. (2011). Data mining: A prediction of performance or underperformer using classification. *International Journal of Computer Science and information technology*, 2(2), pp. 686- 690
- Pittman, K. Comparison of data mining techniques used to predict student retention. PhD thesis, Nova Southeastern University, 2008



- Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2010). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*. Doi: 10.1002/cae.20456
- Romero, C., Ventura, S., Espejo, P. G., & Hervás, C. (2008, June). Data Mining Algorithms to Classify Students. In EDM (pp. 8-17)
- Ruby, J. & David, K. (2014). Predicting the Performance of Students in Higher Education Using Data Mining Classification Algorithms Algorithms – A Case Study. *International Journal for research in Applied Science & Engineering Technology (IJRASET)*. Vol. 2 Issue XI, ISSN: 2321-9653
- Sonali, A., G.N. Pandey and M.D. Tiwari, 2012. Data mining in education: Data classification and decision tree approach. *Int. J. e-Educ. E-Bus. eLearn.*, 2(2): 140-144
- Sree, G. & Ruppa, Ch. (2013). Data Mining: Performance Improvement in Education Sector Using Classification and Clustering Algorathim. *IJIRD*. Vol 2, Issue 7
- Surjeet, K., & Saurabh, P. (2012). Data Mining Application in Enrolment Management: A Case Study. *International Journal Of Computer Application (IJCA)* Vol.41 No.5, pp 1-6
- Surjeet K. Y and Saurabh P (2012). Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification. *World of Computer Science and Information Technology Journal (WCSIT)* ISSN: 2221-0741 Vol. 2, No. 2, 51-56, 2012
- Tarun, I., Gerardo, B. and Tanguilig, B. (2014). Predictive Decision Support System for Licensure Examination Performance through Integration of Multiple Regression and PART Classification Models of Data Mining