academicJournals

Vol. 12(24), pp. 1188-1198, 23 December, 2017
DOI: 10.5897/ERR2017.3266
Article Number: 8CACD3D67014
ISSN 1990-3839
Copyright © 2017
Author(s) retain the copyright of this article
http://www.academicjournals.org/ERR

Educational Research and Reviews

*Full Length Research Paper*

# Predicting effective course conduction strategy using Datamining techniques

**Parkavi A[123]*, K. Lakshmi[123], K.G. Srinivasa[4]**

[1]Research Scholar, Periyar Maniammai University, Thanjavur, India.
[2]Ramaiah Institute of Technology, Bangalore, India.
[3]Periyar Maniammai University, Thanjavur, India.
[4]CBP Government Engineering College, New Delhi, India.

Received 3 May, 2017; Accepted 9 June, 2017

**Data analysis techniques can be used to analyze the pattern of data in different fields. Based on the analysis' results, it is recommended that suggestions be provided to decision making authorities. The data mining techniques can be used in educational domain to improve the outcome of the educational sectors. The authors carried out this research study by devising a mathematical model and tool to determine effective course conduction strategy, for teaching the students regression analysis and goodness of fit test. This helps the faculty to take necessary remedial actions to improve performance in courses by selecting effective course conduction based on the students' performance. The accuracy of prediction in this research is more as it is measured by using Rsquare value which is near to one.**

**Key words:** Data mining, educational data analysis, course conduction strategy, outcome based education.

## INTRODUCTION

Large and increasing amount of data enhance the use of data analysis tools to discover regular and irregular pattern of data. Data mining field is growing up as a discipline in which tools can be used to analyze data, reveal undiscovered knowledge and provide automated decisions in different domains. One of the domains in which data mining is used is education domain, which is called educational data mining.

The main purpose of educational organization is to improve the quality of teaching and learning, making the outcome of education effective and stakeholders satisfied. The main purpose of this research work is to propose efficient statistical and data mining models for improving outcome based education. In educational domain, the need for analysis and prediction of students' performance is increasing.

In this paper, a data model is proposed and tested to prove the effectiveness of their usage in educational domain. As institutions are moving with the main goal of outcome based education, it has a plan to improve outcome based education. This can be fulfilled using recommendation system, which can help the students and faculty by providing valuable suggestions to attain outcome based education. The challenge faced related to this research is: To study and analyze performance of students and determine better course conduction strategy.

The study aims to find the trends in teaching and learning and their outcome. The analysis is done using R programming language where chi square goodness of fit

*Corresponding author. E-mail: parkavi.a@msrit.edu.

test is applied to determine and analyze the effectiveness of active learning strategies, used in courses. After the students go through the active learning of a course and take up the Internal and external examinations, they give us a clear scope for evaluation and comparison of outcome of the active learning strategies.

## LITERATURE REVIEW

Data mining applications are enormously getting used in education domain, to improve the performance of the students. Researchers have scientifically investigated data of educational domain using data mining; thus, the outcome of the education will be improved.

Data mining is useful in providing r ecommendations to improve the businesses by analyzing the business data. In this paper, the authors used statistical and data mining techniques to provide recommendations for improving the outcome of education, by analyzing educational data. The learners' behavioral pattern can be mined using data mining, and different levels of recommendations can be produced using recommender systems.

The recommendation system is designed to provide individualized recommendations to students and to improve their learning effectiveness (Huebner). Predictive assessments tools are used to assess the gap between known and unknown knowledge of students during a professional programme. For this purpose, the predictive assessment tool analyzes students' performance in exams.

Projects are used to provide information to teachers, to let them determine which pedagogical techniques are effective in providing better learning style, during teaching by personalizing learning (West, 2012). Teaching and learning through visualizations improve the engagement of students' learning. Improvements in active, collaborative and student centered learning can be achieved through visualization systems with higher engagement levels in class rooms. That will in turn increase the attainment levels of course outcomes (Laakso et al., n. d.).

Data mining model can be designed to acquire and store data around instructors which are collected from different data sources. Using the data mining model of different classification algorithms, knowledge about instructors can be obtained. For this, different classifiers are trailed and best classification algorithm is found out to get the best prediction results. Then the patterns of instructor data are mapped to generate rules using best classification algorithm to predict instructors' performance and provide recommendation to improve instructors' performance (Ola and Palaniappan, n. d.).

### Problem statement

The authors wish to analyze the performance of the students who have gone through different teaching strategies (Hernandez et al., 2015). The study is aimed at teaching trends followed to deliver same course for two different batches. Analysis of students' performance comparison is done using R programming where chi square goodness of fit test is applied to analyze the impact of different course conduction strategies.

Further, authors did analysis which helps the course conduction faculties to determine which course conduction strategy is better. Here an experimental study is done to analyze impact difference in students' performance under various course conduction strategies. The same data analysis can be used in a scenario where the same faculty has conducted the same course for two different batches of students, using same course conduction strategy.

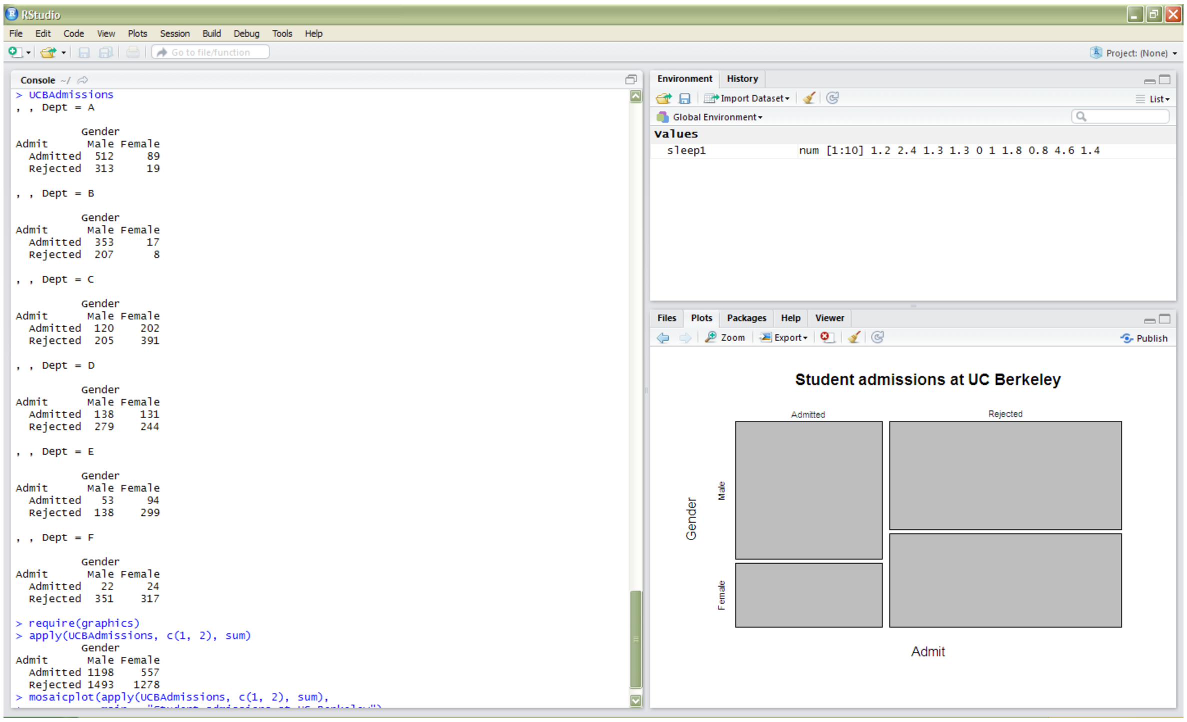
There, to compare the performance of batches or to compare the effectiveness of faculty, the same analysis can be used. If the students' performance is graded down, recommendation has to be provided to faculty about the degradation of students' performance. If two different faculties have used same course conduction strategy for two different batches, then to compare their effectiveness the same analysis can be used. So, based on that the recommendation can be provided to faculty whose students have performed low.

If a faculty is teaching the same course consecutively then her/his effectiveness can be predicted for forthcoming batch. For this, the students' batches performance has to be gathered and analyzed for whom recently the faculty has taught the same course. The analysis and prediction is done using R programming tool. And the authors have designed a tool for this analysis in python, to perform the analysis in user friendly manner and to provide recommendations.

### Proposed method for comparative analysis of course conductions strategies using chi square goodness of fit test and regression analysis

The effectiveness of course conduction strategies can be analyzed primarily based on students' performance in examinations. The education paradigm is shifted from teacher centered model to learner centered model. Students have the responsibility for learning process in actively engaged in upgrading his/her knowledge. The responsibility of teacher in learning process is to guide and facilitate, to make sure the learning goals are attained properly.

The evaluation strategy is not only evaluating the knowledge but including the learning process also. For this, the theoretical courses are accompanied with practical exercises and the feedback received from learners, thus the active learning can be improved. Educators prefer the learning process of theoretical courses with laboratory activities using the process of



```
> UCBAdmissions
, , Dept = A

         Gender
Admit      Male Female
  Admitted  512     89
  Rejected  313     19

, , Dept = B

         Gender
Admit      Male Female
  Admitted  353     17
  Rejected  207      8

, , Dept = C

         Gender
Admit      Male Female
  Admitted  120    202
  Rejected  205    391

, , Dept = D

         Gender
Admit      Male Female
  Admitted  138    131
  Rejected  279    244

, , Dept = E

         Gender
Admit      Male Female
  Admitted   53     94
  Rejected  138    299

, , Dept = F

         Gender
Admit      Male Female
  Admitted   22     24
  Rejected  351    317

> require(graphics)
> apply(UCBAdmissions, c(1, 2), sum)
         Gender
Admit      Male Female
  Admitted 1198    557
  Rejected 1493   1278
> mosaicplot(apply(UCBAdmissions, c(1, 2), sum),
+            main = "Student admissions at UC Berkeley")
```

```
Values
  sleep1             num [1:10] 1.2 2.4 1.3 1.3 0 1 1.8 0.8 4.6 1.4
```

**Figure 1.** R studio workspace.

learning by doing. The learning effectiveness of different strategies can be analyzed by comparing examination results of different batches, where different course conduction strategies are used accompanying various active learning methods in conducting classes.

For that, the students' performance in active learning tasks, continuous internal assessments and semester end examinations can be collected and analyzed to identify better course conduction strategy. The authors have focused on the study by focusing on this aspect using R programming to determine the significance of the different course conduction strategies and to identify the better strategy based on students' performance. These authors have designed a tool for giving recommendations using python.

## R programming

R programming is used for modern statistics. R has more statistical packages. Data can be arranged using variables of columns and rows.

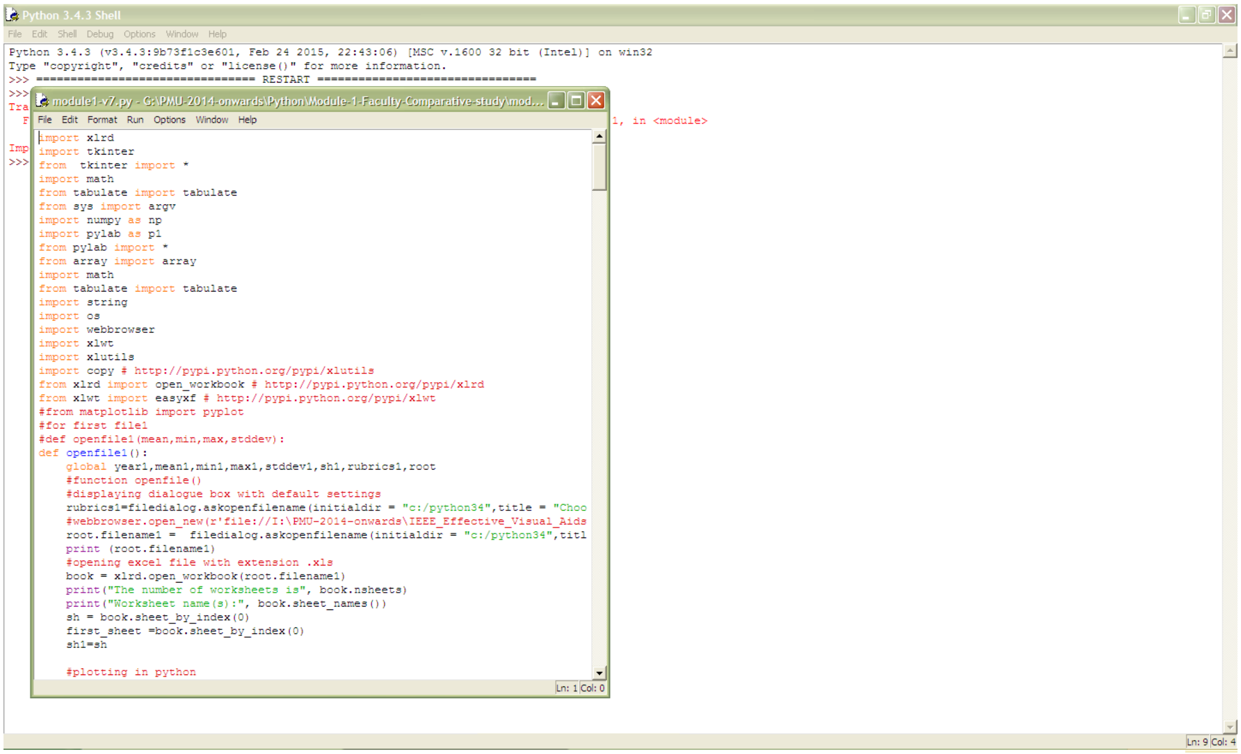
R helps to do different kinds of functions like manipulation, statistical modeling and graphics. Extensibility is a greater advantage of using R. Addition on packages can be easily developed using R programming. R is a good programming language to perform all varieties of statistical analysis for practitioners and researchers.

Using R, the statistical analysis like linear and nonlinear modeling, classical statistical tests, time-series analysis and clustering and classification analysis can be done (The R Project for Statistical Computing, n.d.). R environment provides the following facilities (Why use the R Language?, n.d.): Effectively handling data and storage facility of data, data structures, handling of missing data, suitable operators to perform calculations on arrays, Intermediate data analysis tools collections, graphical representation of data analysis, input, output , conditional, looping and user defined functional statements.

## R studio

R studio is an integrated development environment (IDE) for R programming language. This IDE provides a console and editor supports syntax highlighting. R studio provides facility to debug code, execute code and tools for plotting. This IDE eases the management of workspace. R studio includes workspace browser and data viewer.

This IDE brings our workflow all together with powerful authoring and debugging as shown in **Error! Reference source not found.**. R studio provides facility to integrate the tools, the authors use with R into single environment. Rapid navigation to files and functions can be easily done in R studio. Support for authoring html, pdf, word and slides are available with R Studio. Interactive graphics support is also provided in R Studio (Take control of your R code, n.d.).


Python 3.4.3 Shell
File Edit Shell Debug Options Window Help
module1-v7.py - G:\PMU-2014-onwards\Python\Module-1-Faculty-Comparative-study\mod...
File Edit Format Run Options Window Help
Ln: 1 Col: 0
Ln: 9 Col: 4

```
Python 3.4.3 (v3.4.3:9b73f1c3e601, Feb 24 2015, 22:43:06) [MSC v.1600 32 bit (Intel)] on win32
Type "copyright", "credits" or "license()" for more information.
>>> ================================ RESTART ================================
>>>
Tra
  F                                                              1, in <module>

Imp
>>>
```

```
import xlrd
import tkinter
from  tkinter import *
import math
from tabulate import tabulate
from sys import argv
import numpy as np
import pylab as p1
from pylab import *
from array import array
import math
from tabulate import tabulate
import string
import os
import webbrowser
import xlwt
import xlutils
import copy # http://pypi.python.org/pypi/xlutils
from xlrd import open_workbook # http://pypi.python.org/pypi/xlrd
from xlwt import easyxf # http://pypi.python.org/pypi/xlwt
#from matplotlib import pyplot
#for first file1
#def openfile1(mean,min,max,stddev):
def openfile1():
    global year1,mean1,min1,max1,stddev1,sh1,rubrics1,root
    #function openfile()
    #displaying dialogue box with default settings
    rubrics1=filedialog.askopenfilename(initialdir = "c:/python34",title = "Choo
    #webbrowser.open_new(r'file://I:\PMU-2014-onwards\IEEE_Effective_Visual_Aids
    root.filename1 =  filedialog.askopenfilename(initialdir = "c:/python34",titl
    print (root.filename1)
    #opening excel file with extension .xls
    book = xlrd.open_workbook(root.filename1)
    print("The number of worksheets is", book.nsheets)
    print("Worksheet name(s):", book.sheet_names())
    sh = book.sheet_by_index(0)
    first_sheet =book.sheet_by_index(0)
    sh1=sh

    #plotting in python
```

**Figure 2.** Python workspace.

## Python

Python programming language is a dynamic object oriented programming language. Programmer can integrate many technologies using python, thus the productivity in software development life cycle can be efficient. It is available for most operating systems. It provides extensive support libraries and provides facility to work with different data formats. Numeric and scientific applications can be developed using numeric python and scientific python (Benefits of Python, n.d.).

Python can be used as scripting language to customize or extend the applications as shown in Figure 2. Python packages are useful to develop good quality graphical user interfaced applications as shown in Figure 3. This language has agile nature so the rapid development of prototype is easy. Python is an open source, so contributors from all over the world are working on improvising the designing and extending the features of python.

## Problem formulation

### Study objective

The comparative analysis of different strategies followed in conducting course helps the teaching faculty to draw some useful conclusions and decisions. In this study, the authors have made an experimental attempt to analyze the variations in trends for conducting courses.

## Implementation details for predicting better course conduction strategy

The predictive frequency is obtained in the following equation:

PFR=TSRS*TSPD/TSAD where
PFR is Predictive frequency in particular mark range;
TSRS is Total number of students in different mark ranges under a single course conduction strategy;
TSPD is Total number of students in a particular range under different course conduction strategy;
TSAD is Total number of students in all mark ranges under different course conduction strategy;

## System architecture

Consider S is the system which describes student dataset that is, set of data items with students' performance details in same course under different course conduction strategies.

This course level performance data set is the input to our system for statistical analysis and goodness of fit test to identify the better course condition strategy for a

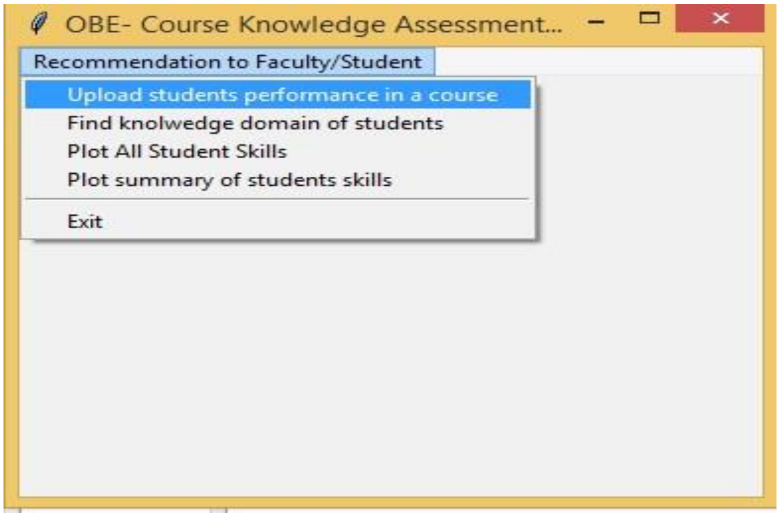

OBE- Course Knowledge Assessment...
Recommendation to Faculty/Student
Upload students performance in a course
Find knolwedge domain of students
Plot All Student Skills
Plot summary of students skills
Exit

**Figure 3.** GUI developed using python.

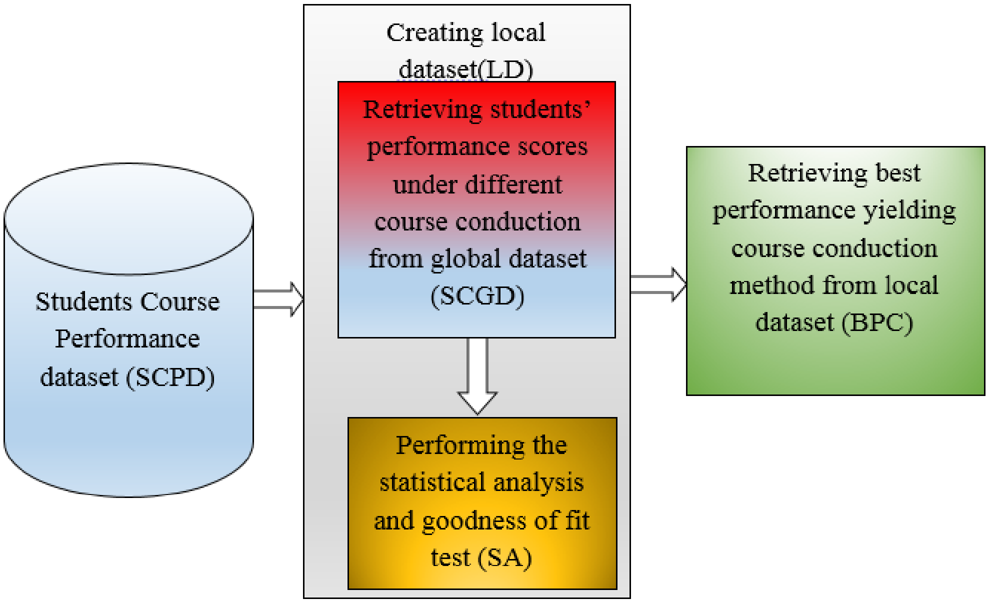

Students Course Performance dataset (SCPD)
Creating local dataset(LD)
Retrieving students' performance scores under different course conduction from global dataset (SCGD)
Performing the statistical analysis and goodness of fit test (SA)
Retrieving best performance yielding course conduction method from local dataset (BPC)

**Figure 4.** System for predicting better course conduction strategy for a course mathematical model.

course. Then creating the local data sets is done to predict the better course conduction strategy (Figure 4)

### Variables

S=(SCPD,LD,SCGD,SA,BPC)
SCPD=Students courses Performance dataset;
LD=Creating local dataset;
SCGD=Retrieving students' performance scores under different course conductions from global dataset;
SA= Performing the statistical analysis and goodness of fit test;
BPC= Retrieving best performance yielding course conduction method from local dataset;
CCSA=Course conduction strategy analyzer;
SCP= Students' course performance details.

### Inputs

SCP={$PD_1$,$PD_2$,SD}

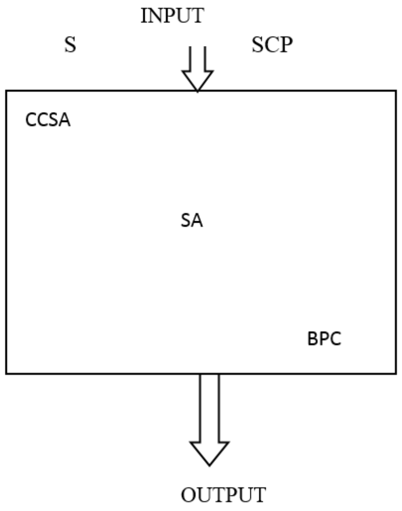


**Figure 1.** Mathematical Model for predicting better course conduction strategy for a course approach

Where students courses performance dataset, $PD_1=\{P_{11},P_{12},P_{13},\ldots P_{1N1}\}$ is a set of items and for each performance item $P_{1id1}$ ($1<=id1<=N1$) has a unique id, called $P_{id1}$ where 'N1' is the number of students offering a course using course conduction strategy method-1.
$PD_2=\{P_{21},P_{22},P_{23},\ldots P_{2N2}\}$ is a set of items and for each performance item $P2_{id2}$ ($1<=id2<=N2$) has a unique id, called $P_{id2}$ where 'N2' is the number of students doing a course using course conduction strategy method-2.
$P_{ji}=s$ where s is the marks scored by i-th student in a course and j =id1 or j=id2
Statistical Data =SD={SD1,SD2}
SD1= {mean, standard deviation, max, min} of PD1
SD2={mean, standard deviation, max, min} of PD2

## Process

$$\text{CCSA}= \bigcup_{i=1}^{2} [PD_i = \bigcup_{k=1}^{n_i} P_{ki}] \quad (1)$$

$$\text{SA}= \bigcup_{i=1}^{2} [SD_i = \{\text{mean}(PD_i),\text{stddev}(PD_i),\max(PD_i),\min(PD_i)\}] \quad (2)$$

(3) BPC = max (SA1,SA2) (Figure 5).

The course conduction strategies can be compared with respect to the students' performance in internal tests and semester end exams. Analysis of students' performance attained under different course teaching and learning strategies makes the faculty to follow or improvise certain course conduction methodology.

To identify the significance of various course conduction methodologies, the authors find out the students' performance in different mark ranges. Then, the authors performed chi square goodness of fit test to analyze whether the different teaching learning process made an impact over the students' performance or not.

For this study, during the last three years, the authors used different teaching and learning strategies for the following courses: Data structure with C, System software and Compiler design. The authors used different active learning and collaborative activities to improve students' learning. Then the authors collected students' performance under different course conduction for same course and did the comparative analysis. The activities carried out for this study consist of the following steps:

(1) Collection of data
(2) Preparation of data
(3) Preprocessing the data
(4) Processing the data
(5) Results and analysis

In this study, the authors considered an engineering organization, MS Ramaiah institute of Technology. The authors made use of data set in ".csv" format. It contains a batch of 154 students' performance in Continuous Internal Evaluation for a course. For instance, they consider the years 2013 and 2014. A sample of the dataset prepared for this study is shown in Figure 6 (Table 1).

## Collection and preparation of data

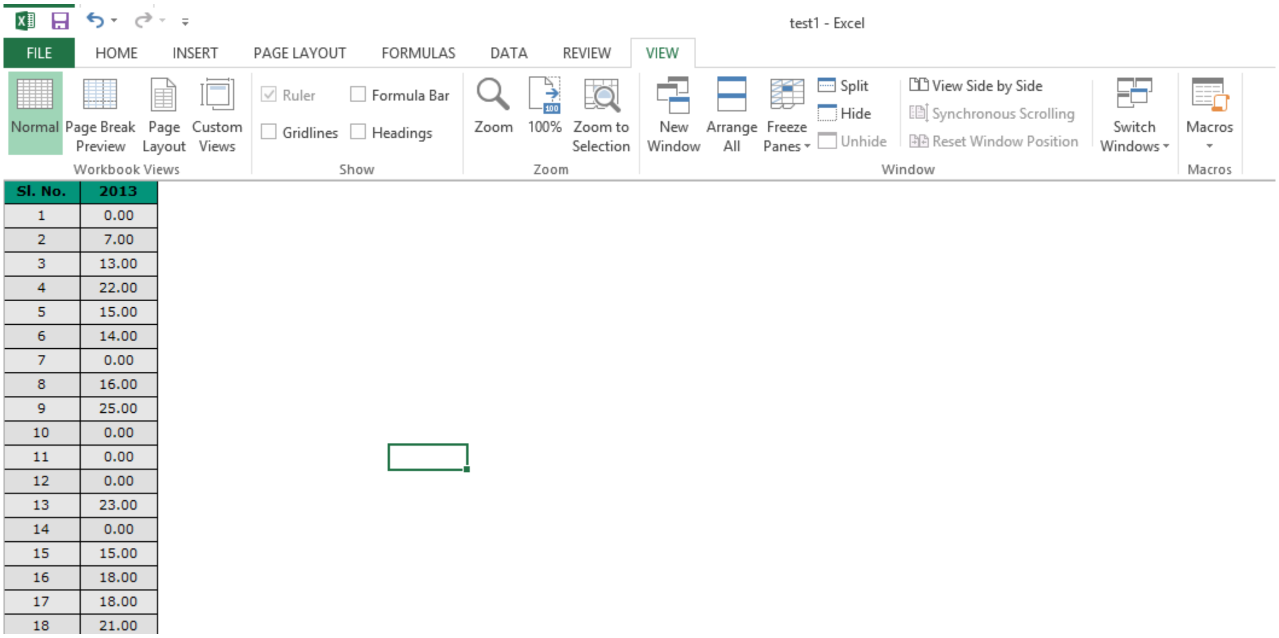
### *Preprocessing of data*

A faculty, who wants to take decision about course conduction strategies, analyzes the historical data of students' performance, which are obtained from Institution's examination repository. In this stage data are identified, gathered, cleaned and aggregated into a format needed for our data models. Here, the missing values are filled with "A" to represent absent.

## Processing data (Mining the data)

Data mining technique – Chi square goodness of fit test is applied here to analyze whether the course conduction methodology has influenced the students' performance in internal and external exams or not. Chi square goodness of test is one of the oldest and most well-known methods of statistical analysis. It is the process of statistical test used to compare observed data with the data.

The authors expect to obtain with respect to a specific hypothesis. By measuring the deviations which are the differences between observed and expected, the authors can conclude that something other than chance is at work, making the observed to vary from expected. The chi square goodness of fit test is called testing null



| Sl. No. | 2013 |
|---|---|
| 1 | 0.00 |
| 2 | 7.00 |
| 3 | 13.00 |
| 4 | 22.00 |
| 5 | 15.00 |
| 6 | 14.00 |
| 7 | 0.00 |
| 8 | 16.00 |
| 9 | 25.00 |
| 10 | 0.00 |
| 11 | 0.00 |
| 12 | 0.00 |
| 13 | 23.00 |
| 14 | 0.00 |
| 15 | 15.00 |
| 16 | 18.00 |
| 17 | 18.00 |
| 18 | 21.00 |

**Figure 2.** Sample data set.

**Table 1.** Structure of data.

| S/N | Attribute name | Type | Description |
|---|---|---|---|
| 1 | Roll No | Number | Roll Number of student |
| 2 | Marks | Number | Marks scored by student |

hypothesis which states that between the expected and observed results there are no significant results. In R studio the authors can work with any number of attributes. There are various packages which help us to perform useful statistical analysis over our data sets to determine the attribute nature of dependency or independency.

The data are configured into the ‘Comma Separated Values’ using ‘Microsoft Excel’ node. The students’ performance is accepted using R script. Students in different ranges are measured for two different batches of two different course conduction strategies. Then expected frequencies are calculated to find variance from observed frequencies of students’ performance to measure chi square value. The p-value significance level is used to determine whether the students’ performance in different ranges of different batches is independent of course conduction strategy or not.

### *Analyzing students’ performance*

The dependency nature of students’ performance in different ranges is analyzed to obtain useful results as seen in the next section.

## Results and discussion for comparing course conduction strategies using chi square goodness of fit test

### *Reports*

In this study, the authors get many reports which are valuable assets. The chi square goodness of fit test begins with measurement of predictive frequencies. The predictive frequencies are formed based on students’ performance in different ranges under different course conduction strategies. After the measurement of predictive frequencies, a variety of measures of statistical significance such as chi square and significance level of p-value are included. The predictive frequency is obtained in the following equation:

PFR=TSRS*TSPD/TSAD where;
PFR is Predictive frequency in particular mark range; TSRS is Total number of students in different mark ranges under a single course conduction strategy; TSPD

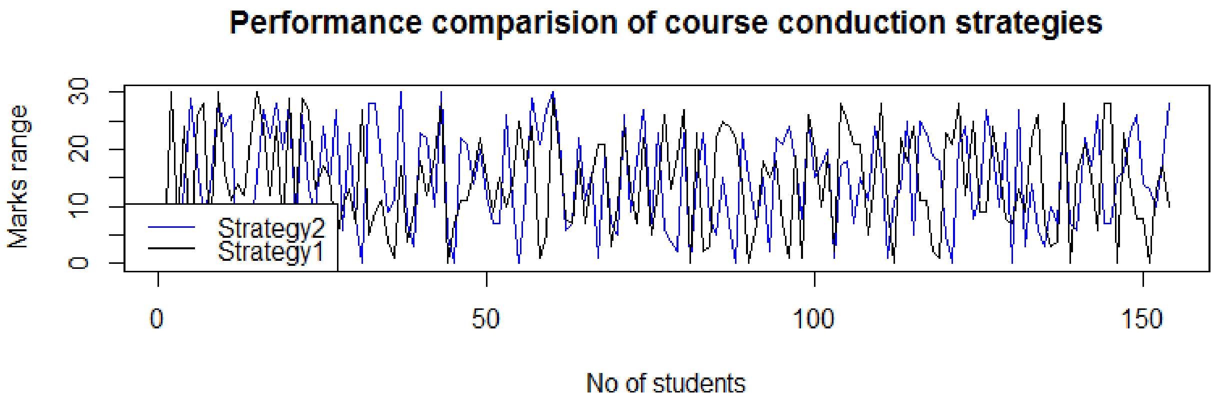
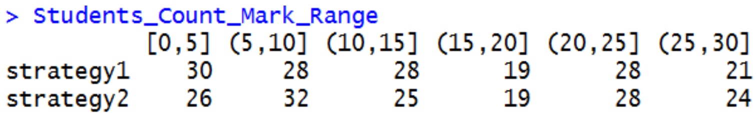


**Figure 3.** Performance comparison of students.



```
> Students_Count_Mark_Range
          [0,5] (5,10] (10,15] (15,20] (20,25] (25,30]
strategy1    30     28      28      19      28      21
strategy2    26     32      25      19      28      24
```

**Figure 8.** Students' count in different marks range.

is Total number of students in a particular range under different course conduction strategy; TSAD is Total number of students in all mark ranges under different course conduction strategy.

Based on this aforementioned equation, predictive frequencies measured for students' performance in different marks ranges are obtained. The variations of these predicted frequencies and observed (original) frequencies are shown in the form of R programming graphs (Figure 8).

### *Performance comparison of students using basic statistical analysis in Python*

For performance, comparison of two different course conduction strategies, basic statistical analysis is performed over the performance data sets of students. The comparative analysis of students performance is shown in Figure 7a.

That is why we get the line graphs which are similar, without much variation, and show the performance of students achieved by two different course conduction methods are similar. Number of students are shown in x axis and the students marks are shown in Y axis.

The numbers of students using strategy 1 and 2 in different marks ranges are shown in Figure 8. The number of students scored between 0 to 5 mark range under course conductions trategy1 is 30; with strategy2 it is 26. Similarly, for other ranges also the comparison is shown in the same Figure 8.

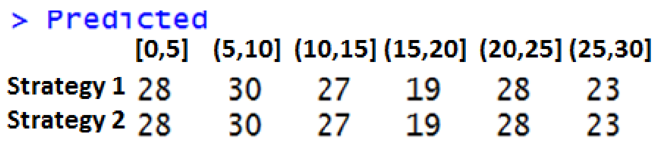
In Figure 9, the predicted students count for each marks range is shown for strategy 1 and 2. As this prediction is based on the past history performance obtained through strategy1 and 2, the predicted results for both strategy are same.

### Performance comparison of students under different strategies using chi square goodness of fit test in R programming

For chi square goodness of fit test, predicted frequencies are measured based on students' performance of two different batches and two different course conduction strategies. The comparative analysis of students' performance of a course conductions strategy and predicted performance are shown in Figure 10. The performance comparison graph shows the data points from original dataset that is explored along with predicted data points using chi square goodness of fit test. This helps us to get a quick view over accuracy and predictive nature of model. The data point locations are shown along the x-axis with respect to mark range in y-axis (Figure 11).

### *Linear regression analysis of students' performance*

Students' performances are analyzed using linear regression analysis with respect to original observed



```
> Predicted
           [0,5]  (5,10]  (10,15] (15,20] (20,25] (25,30]
Strategy 1  28     30      27      19      28      23
Strategy 2  28     30      27      19      28      23
```

**Figure 9.** Students predicted count in different marks range using strategy1 and 2.

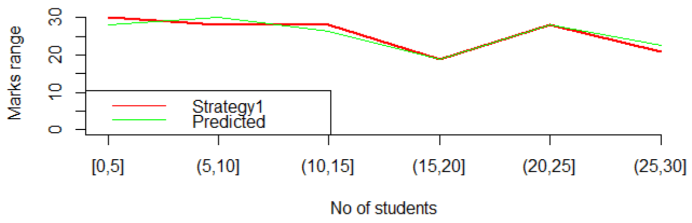


**Figure 10.** Performance comparison of students' performance under course conduction strategy1 and predicted performance.

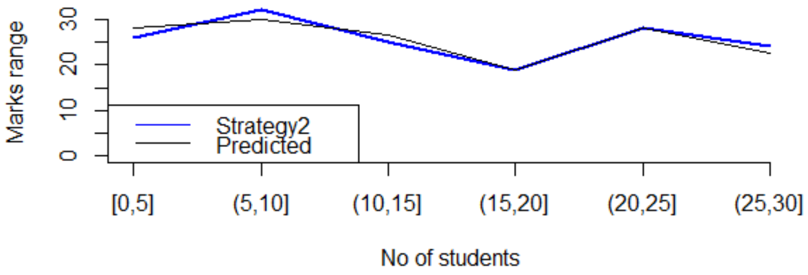


**Figure 11.** Performance comparison of students' performance under course conduction strategy 2 and predicted performance.

performances and the predicted performances.

### *Performance analysis of students using Linear Regression analysis*

The predicted performance versus original performance of students using course conduction strategy1 shows all the performance data points of original dataset that are explored along with performance data points that would have been predicted using the model produced. This helps us to have overall view about the accuracy and predictive power of model. The original performance dataset of strategy1 is shown along x-axis and the predicted performance data set is shown along the y-axis (Figure 12).

Similarly, the predicted performance vs original performance of students using course conduction strategy2 can also be analyzed using Linear Regression analysis model as shown in Figure 13.

### *Prediction accuracy*

The R-Square values are used to determine the accuracy of predicted students' performance. If the R square value is nearer to '1' it represents that the prediction is accurate. The R square value obtained for the predictions of two strategies performances are 0.80 and 0.73. The variations occur to some extent due to the institutions' intake of candidates' knowledge level (R tutorial, 2016). The performance using strategy1 vs strategy2 of student can also be analyzed using Linear Regression analysis model as shown in Figure 14.

## Conclusion

In this study, the authors carried out this research study

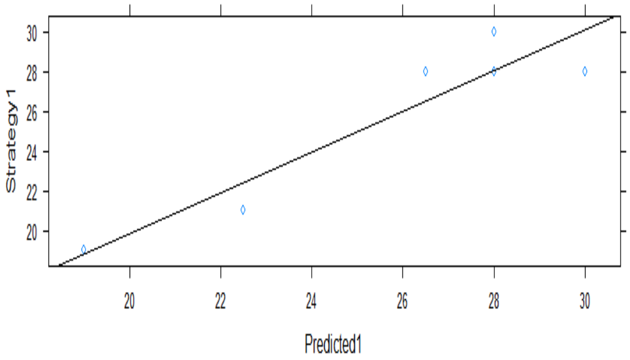


**Figure 12.** Predicted performance vs performance of students under course conduction strategy1 using Linear Regression.

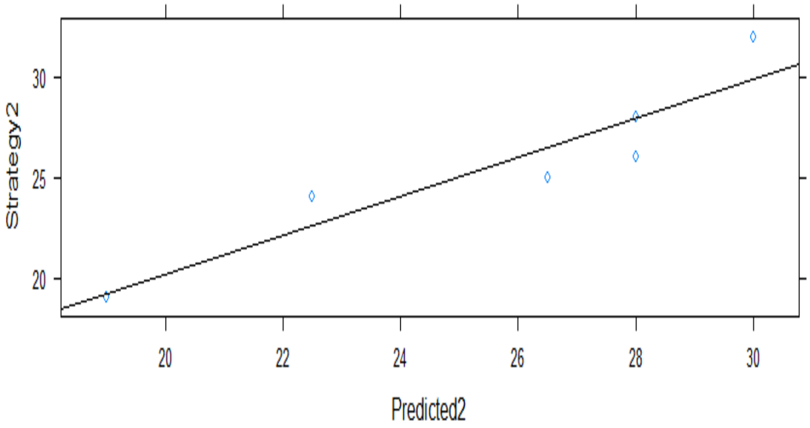


**Figure 13.** Predicted performance vs performance of students under course conduction strategy2 using linear regression.

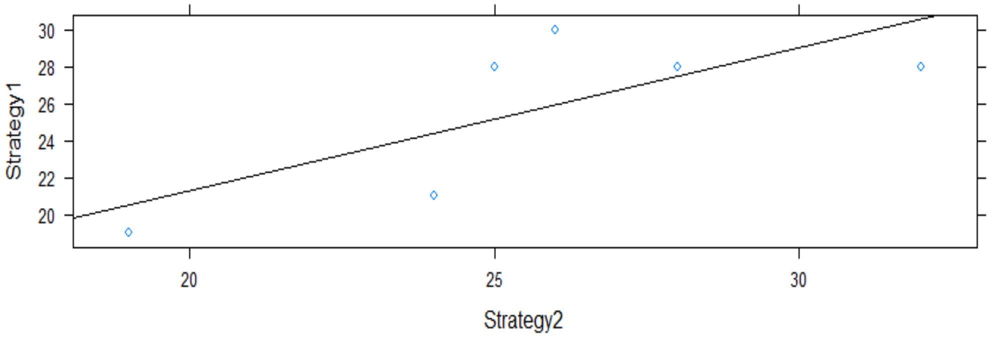


**Figure 14.** Performance of students under course conduction strategy1 vs strategy2 using linear regression.

to determine the effective course conduction strategy for a course based on students' performance using regression analysis and goodness of fit test. This helps to identify the better course conduction strategy by comparative analysis of students’ performance in internal and external examinations. Based on the analysis, results remedial action can be taken to improve students’ performance by updating or improvising the course conduction strategies.

## CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

### REFERENCES

Benefits of Python. (n.d.). (Digital Mesh Softech India (P) Limited, Kochi ) Retrieved from http://www.digitalmesh.com/offshore-development-center/python-development/benefits-of-python.html

Hernandez-Jayo U, Juan-Manuel López-Garde, Rodríguez-Seco J (Nov, 2015). Addressing Electronic Communications System Learning Through a Radar-Based Active Learning Project. IEEE TRANSACTIONS ON EDUCATION, 58(4):269-275.

Laakso MJ, Myller N, Korhonen A (n.d.). Comparing Learning Performance of Students Using Algorithm Visualizations. Educ. Technol. Society, 12(2):267–282.

Ola AF, Palaniappan S (n.d.). Design of an Intelligent Algorithm for Evaluation of Instructors’ Performance in Higher Institutions of Learning using Data Mining. Academic Research Online Publisher.

Tutorial R (2016). Retrieved from http://www.r-tutor.com/elementary-statistics/simple-linear-regression/coefficient-determination

Take control of your R code. (n.d.). (R Studio) Retrieved from https://www.rstudio.com/products/RStudio/

The R Project for Statistical Computing. (n.d.). (The R Foundation) Retrieved from https://www.r-project.org/about.html

West DM (SEP 2012). Big Data for Education: Data Mining, Data Analytics, and Web Dashboards. (Governance Studies at Brookings)

Why use the R Language? (n.d.) (Burns Statistics) Retrieved from http://www.burns-stat.com/documents/tutorials/why-use-the-r-language/.