Evaluation of Students Performances in an Examination With Data Mining Techniques

Olagunju, M (PhD)
Dept. of Computer Science, IICT
Kwara State Polytechnic
Ilorin, Nigeria
(E-mail-Olamukalla@yahoo.com

Ameen, A. O (PhD)
Dept. of Computer Science
University of Ilorin
Ilorin, Nigeria
E-mail-aminamed@unilorin.edu.ng

Aiyesimi Y.M.(PhD)
Dept. of Mathematics,
Federal University of Technology
Minna, Nigeria
yomiaiyesimi20131959@gmail.com

ABSTRACT

This paper developed an Algorithm to classify students’ data based on their performance in an examination. The C program was used to develop and implement the classification Algorithm. The classification process assists to determine both the truth pass and truth failure rate in a particular examination. The implementation shows promising efficiency in the domain of application.

Keywords - Classification confusion matrix, Algorithm, Truth Pass Rate (TPR), Truth Failure Rate (TFR), Misclassification Rate (MCR), unsupervised learning, supervised learning, Data Mining

1. INTRODUCTION

Data Mining form certain part of Knowledge Database Discovery (KDD) which involve the uses of Algorithm to find the hidden information in the data base (Dunham, 2003). Also data mining can be defined as a process of using computer to analyse large database to determine the needed information (David, 2005). The KDD which the data mining form part of it has the following stages.

1. Selection of data (stage 1)
2. Pre-processing of data (stage 2)
3. Transformation of data (stage 3)
4. Mining of data (stage 4)
5. Interpretation of data (Stage 5)
The KDD stages mentioned above according to Peter et al (1997) can be described briefly as follows:

1. **Selection of data**: This is the first stage in KDD process and it involves the selection of data from the primary source such as selection from Hard copy e.g electronic file
2. **Pre-Processing Stage**: This is the second stage in carrying out KDD activities, it deal with re-structuring of data to meet the need of the user.
3. **Transformation of data**: This is the third stage and is deal with fill margin of data to suit the need of the user.
4. **Data Mining**: This is the fourth stage and it involves finding hidden information in the database.
5. **Interpretation**: It is the last stage of KDD process and it involves the display of output data especially through the visualization process.

1.1 Data Mining in Perspective

Data Mining according has a lot of applications and uses which include:

1. **Terrorists detection**: It is used to detect criminal activities.
2. **Bank use**: It is used in banking systems to carryout financial activities.
3. **Medical Use**: It is used in medicine to diagnose patients.

1.2 Problem in View

Data Mining is utilized in this research to classify students’ results based on available data in order to optimize the accuracy of their classification (Ian & Ebie, 2000).

2. THE DATA MINING PROCESS

There are different types of Data Mining process available (David, 2005; Richard et al, 1980). In this paper, the cross industries standard process mining simply refer to as CRS-DM) shall be the focus. It involved the following stages.

1. **Business objective**: This deals with the definition of the objectives involved in carrying out mining activities
2. **Data Understanding** This is the second stage of the data mining process and it involves the under study of data to be used in data mining process (Trevor et al, 2001).
3. **Data Preparation**: This is the third stage it involves cleaning of data to remove unwanted items from the data available.
4. **Modeling**: It deals with the development of the model with any chosen programming languages.
5. **Testing of Data**: It involves the testing of the model that is developed in the previous stage
6. **Evaluation**: it deals with the evaluation of the modeling process.
7. **Documentation**: it involves the documentation of the previous stage and that is possible if the model that is developed is tested.
The diagrammatical representation of data mining process can be given below according to Yike and Robert (1997).

3. MATERIALS AND METHODS

The materials used are the student results shown in the table 1.2 and 1.3 respectively. The method used in the development of algorithm is given below the C program was developed to assist in the determination of TPR, TFR, MCR and AR.

Algorithm
1. Start
2. Input grade
   If (grade ≥ 0.0) and (grade ≥ 1.9a)
      Class = fail
   Else
      If (grade ≥ 2.0) and (grade ≤ 2.49)
         Class = lower
      Else
         If (grade ≥ 2.50) and (grade ≤ 3.49)
            Class = lower
         Else
            If (grade ≥ 3.5) and (grade ≤ 3.49)
               Class = Upper
   Else
5. Stop
The above algorithm can be interpreted in the confusion matrix as shown below: in Table 1: Confusion Matrix

<table>
<thead>
<tr>
<th></th>
<th>PASS</th>
<th>FAIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
<td>TP</td>
<td>TF</td>
</tr>
<tr>
<td>FALSE</td>
<td>FP</td>
<td>FF</td>
</tr>
</tbody>
</table>

The above confusion matrix (2 x 2 matrix) according to Trevor & Robert (2009) can be need to represent the classification output of the based algorithm developed can be related to inputs of confusion matrix as shown below:

TP = Class as upper or distinction
TF = Fail or withdrawal
FP = Lower
FF = Pass

The above terms can be interpreted as follow:
TP = truly classified as Pass
TF = truly classified as fail
FP = false fully classified as fail

The TPR and TFR and MCR and Accuracy rates can be calculated by the questions given below.

\[ \text{TPR} = \frac{TP}{TP + FF} \times 100 = \ldots \ldots \ldots \ldots \text{equation 1} \]

\[ \text{TFR} = \frac{TF}{TP + TF} \times 100 = \ldots \ldots \ldots \ldots \text{equation 2} \]

\[ \text{MCR} = \frac{FP + FF}{TP + TF + FP + FF} \times 100 \ldots \ldots \ldots \text{equation 3} \]

Where

Accurate classificate rate = 1 – MCR

Experiment 1

The formular given above can now used to determine the TPR, TFR and MCR as follow with date in figure 1.2

TR = 2
FF = 16
TF = 1
FP = 6

\[ \text{TPR} = \frac{TP}{TP + FF} \times 100 = \ldots \ldots \ldots \ldots \text{equation 2} \]

\[ \text{TP} = \frac{2}{TP + FF} \times 100 \]

\[ = 2 \times 100 = 11.11\% \]
**Experiment**
The table 1.3 can also be used to calculate TPR, TFR, MCR and accuracy as follow;

\[
\text{TPR} = \frac{\text{TP} \times 100}{\text{TP} + \text{FF}} = \frac{3 \times 100}{3 + 9} = \frac{300}{12} = 25% \\
\text{TFP} = \frac{\text{TF} \times 100}{\text{TF} + \text{FF}} = \frac{2 \times 100}{2 + 13} = \frac{200}{15} = 13.3%
\]

\[
\text{MCR} = \frac{\text{FP} + \text{FF}}{\text{FP} + \text{FF} + \text{TP} + \text{TF}} = \frac{13 + 9}{27} = \frac{22}{27} = 81.5
\]

\[
\text{Accuracy} = 100 \% - 81.5 = 18.5\
\]

**Table 2: NDIII Business Studies Result**

KWARA STATE POLYTECHNIC, ILORIN
CENTRE FOR CONTINUIN EDUCATION
DEPARTMENT OF BUSINESS ADMINISTRATION

<table>
<thead>
<tr>
<th>RESULTS</th>
<th>GRADE</th>
<th>GRADE POINT</th>
<th>RANGE</th>
<th>C.G.P.A</th>
<th>CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMMARY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL NO. OF STUDENTS</td>
<td>95</td>
<td>A</td>
<td>4.00</td>
<td>75% 100%</td>
<td>3.50 - 4.00</td>
</tr>
<tr>
<td>TOTAL NO. WITH DISTINCTION</td>
<td>0</td>
<td>AB</td>
<td>3.50</td>
<td>70% 74%</td>
<td>3.00 - 3.49</td>
</tr>
<tr>
<td>TOTAL NO. WITH UPPER CREDIT</td>
<td>2</td>
<td>B</td>
<td>3.25</td>
<td>65% 69%</td>
<td>2.50 - 2.99</td>
</tr>
<tr>
<td>TOTAL NO. WITH LOWER CREDIT</td>
<td>76</td>
<td>BC</td>
<td>2.75</td>
<td>55% 59%</td>
<td>2.00 - 2.49</td>
</tr>
<tr>
<td>TOTAL NO. WITH PASS</td>
<td>16</td>
<td>C</td>
<td>2.50</td>
<td>50% 54%</td>
<td>Below 2.00</td>
</tr>
<tr>
<td>TOTAL NO. WITH C'OVER</td>
<td>1</td>
<td>CD</td>
<td>2.50</td>
<td>50% 54%</td>
<td></td>
</tr>
<tr>
<td>TOTAL NO. TO REPEAT</td>
<td>0</td>
<td>D</td>
<td>2.25</td>
<td>45% 49%</td>
<td></td>
</tr>
<tr>
<td>TOTAL NO. TO WITHDRAW</td>
<td>0</td>
<td>E</td>
<td>2.00</td>
<td>40% 44%</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: ND III Secretariat Studies Result
KWARA STATE POLYTECHNIC, ILORIN
CENTRE FOR CONTINUIN EDUCATION
DEPARTMENT OF BUSINESS ADMINISTRATION
16/12/08  2006/2007 DIPLOMA/CERTIFICATE RESULT
15:08  ND III SECRETARIAT STUDIES (PART TIME)

<table>
<thead>
<tr>
<th>DIPLOMA/CERTIFICATE</th>
<th>RESULTS</th>
<th>GRADE</th>
<th>GRADE POINT</th>
<th>RANGE</th>
<th>C.G.P.A</th>
<th>CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMMARY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL NO. OF STUDENTS = 27</td>
<td>A</td>
<td>4.00</td>
<td>75% 100%</td>
<td>3.50 – 4.00</td>
<td>DISTINCTION</td>
<td></td>
</tr>
<tr>
<td>TOTAL NO. WITH DISTINCTION = 0</td>
<td>B</td>
<td>3.00</td>
<td>60% 74%</td>
<td>3.00 – 3.49</td>
<td>UPPER CREDIT</td>
<td></td>
</tr>
<tr>
<td>TOTAL NO. WITH UPPPER CREDIT = 3</td>
<td>D</td>
<td>2.50</td>
<td>50% 59%</td>
<td>2.50 – 2.99</td>
<td>LOWER CREDIT</td>
<td></td>
</tr>
<tr>
<td>TOTAL NO. WITH LOWER CREDIT = 9</td>
<td>D</td>
<td>2.00</td>
<td>40% 49%</td>
<td>2.00 – 2.49</td>
<td>PASS</td>
<td></td>
</tr>
<tr>
<td>TOTAL NO. WITH PASS = 13</td>
<td>F</td>
<td>0.00</td>
<td>00% 39%</td>
<td>Below 2.00</td>
<td>CAN’T</td>
<td></td>
</tr>
<tr>
<td>TOTAL NO. WITH C’OVER = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GRADUATE</td>
</tr>
<tr>
<td>TOTAL NO. TO REPEAT = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL NO. TO WITHDRAW = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. CONCLUSION, FINDING AND RECOMMENDATION

In this paper, an algorithm was developed and implemented to assist in determining the level of performance of students’ in an examination, It equally measures the level of the instructors performance as well as the nature of the examination.

5. FUTURE WORK

Future work will explore other areas to which the algorithmic framework can be applied.

REFERENCES

APPENDIX

IMPLEMENTATION CODE LISTING

The C program is developed to assist in calculating different TDR, TFR, MCR and AR. The source and object codes are given below:

Code Listing
#include <stdio.h>
#include <conio.h>
#include <iostream.h>
#include <stdlib.h>
#include <string.h>
#include <math.h>

void mainmenu();
void tprModule();
void tfrModule();
void mcrModule();

float tpf;   // TP for Full Time
float fff;   // FF for Full Time
float tpp;   // TP for Part Time
float ffp;   // FP for Full Time

float tprf; //True Positive Rate for Full Time
float tprp; //True Positive Rate for Part Time

float tff; //TF for Full Time
float fpf; //FP for Full Time
float tfp; //TF for Part ime
float fpp; //FP for Part Time

float tfrf; //True Failure Rate for Full Time
float tfrp; //True Failure Rate for Part Time

float mcrf; //MCR For Full Time
float mcrp; //MCR for Part Time

void main(){
    mainmenu();
}

void mainmenu() {
    int option;
    clrscr();
    cout << "n";
    cout << "n";
    cout << "n";
    cout << "n";
    cout << "n";
    cout << 1. Get True Positive Rate |
    cout << 2. Get True Failure Rate |
    cout << 3. Get Marginal Misclassification Rate |
}
cout << "n               | 4. EXIT                     |
cout << "n               |                            |
cout << "n               | Select from Options (1 - 4) above: |
cout << "n               |______________________________________________ |

gotoxy (56,11); cin >> option;

switch(option) {
  case 1: tprModule(); break;
  case 2: tfrModule(); break;
  case 3: mcrModule(); break;
  case 4: exit(2); break;
  default: {
    cout << "n Invalid Input";
  }
  break;
}

void tprModule () {
  char q;
  q='N';
  clrscr();

  //Accept variables for Full Time
  printf("Enter the True Positive for Full Time: ");
  scanf("%f",&tpf);
  printf("Enter the False Fail for Full Time: ");
  scanf("%f",&fff);

  //Accept Variables for Part Time
  printf("Enter the True Positive for Part Time: ");
  scanf("%f",&tpp);
  printf("Enter the False Fail for Part Time: ");
  scanf("%f",&ffp);

  //Calculate TPR for Full Time
  tprf = (tpf * 100) / (tpf + fff);

  //Calculate TPR for Part Time
  tprp = (tpp * 100) / (tpp + ffp);

  //Display the Result for Full Time and Part Time
  printf("n
\n\n\nTrue Positive Rate for Full Time is ",tprf,"% ");
  printf("True Positive Rate for Part Time is ",tprp,"% ");

cout <<"n\n\nGo Back to Mainmenu? Y/N:"
if (q=='y' || q=='Y') {
  clrscr();
  mainmenu();
}
void tfrModule () {
    char q;
    q='N';
    clrscr();

    //Accept variables for Full Time
    printf("Enter the True Failure for Full Time: ");
    scanf("%f", &tff);
    printf("Enter the False Positive for Full Time: ");
    scanf("%f", &fpf);

    //Accept Variables for Part Time
    printf("Enter the True Failure for Part Time: ");
    scanf("%f", &tfp);
    printf("Enter the False Positive for Part Time: ");
    scanf("%f", &fpp);

    //Calculate TFR for Full Time
    tfrf = (tff * 100) / (tff + fpf);

    //Calculate TFR for Part Time
    tfrp = (tfp * 100) / (tfp + fpp);

    //Display the Result for Full Time and Part Time
    printf("\n\n\n\n%sf%f%s\n", "True Failure Rate for Full Time is ", tfrf, "%");
    printf("%sf%f%s\n", "True Failure Rate for Part Time is ", tfrp, "%");

    cout << "\nGo Back to Mainmenu? Y/N: " << q;
    if (q=='y' || q=='Y') {
        clrscr();
        mainmenu();
    }

    }
}
```c
void mcrModule () {
    char q;
    q='N';
    clrscr();
    // Accept MCR Variables for Full Time
    printf("Enter the False Positive for Full Time: ");
    scanf("%f",&fpf);
    printf("Enter the False Fail for Full Time: ");
    scanf("%f",&fff);
    printf("Enter the True Positive for Full Time: ");
    scanf("%f",&tpf);
    printf("Enter the True Failure for Full Time: ");
    scanf("%f",&tff);
    // Accept MCR Variables for Part Time
    printf("Enter the False Positive for Part Time: ");
    scanf("%f",&fpp);
    printf("Enter the False Fail for Part Time: ");
    scanf("%f",&ffp);
    printf("Enter the True Positive for Part Time: ");
    scanf("%f",&tpp);
    printf("Enter the True Failure for Part Time: ");
    scanf("%f",&tfp);
    //Calculate MCR for Full Time
    mcrf = (fpf * fff * 100) / (tpf + tff + fpf + fff);
    //Calculate MCR For Part Time
    mcrp = (fpp * ffp * 100) / (tpp + tfp + fpp + ffp);
    //Display MCR result for Both Full Time and Part Time
    printf("\n\n\n%sn%f%s\n","Marginal Misclassification Rate for Full Time is ",mcrf,"%n");
    printf("%sn%f%s\n","Marginal Misclassification Rate for Part Time is ",mcrp,"%n");
    cout <<"\n\nGo Back to Mainmenu? Y/N: "; cin >> q;
    if (q=='Y' || q=='y') {
        clrscr();
        mainmenu();
    }
}
}```