

EVALUATION OF STUDENTS PERFORMANCES IN AN EXAMINATION WITH DATA MINING TECHNIQUES

Olagunju, Mukaila

Dept. of Computer Science,
Kwara State Polytechnic
Ilorin, Kwara State, Nigeria
E-mail-Olamukaila@yahoo.com

ABSTRACT

This paper developed an Algorithm to classified the students data based on their performance in an examination. The C program was developed to carry-out the classification used based on Algorithm developed. The classification process assists to determine both the truth pass and truth failure rate in a particular examination.

Keywords and phrases:- Classification confusion matrix, Algorithm, Truth Pass Rate (TPR), Truth Failure Rate (TFR), Misclassification rate (MCR), unsupervised learning, supervised learning, Data Mining e.t.c.

1. INTRODUCTION

Data Mining according to [3], form certain part of KDD (knowledge database discovery) which involve the uses of Algorithm to find the hidden information in the data base [1]. Also data mining can be according to Zanari defined as a process of using computer to analyse large database to determine the needed information. The KDD which the data mining form part of it has the following according to [3] stages or parts.

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)

With reference to [3], [5],

The KDD stages mentioned above according to [3] can be described briefly as follow:

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 fourth stage and it involve find of hidden information in the data base.
5. **Interpretation:-** is the last stage of KDD process and it involves the displaying of output data especially through the visualization process.

Why data mining?

Data Mining according to [2] 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 system to carryout financial activities.
3. **Medical Use:** it is used in medicine to diagnose the patients based on the reward available.

2. STATEMENTS OF THE PROBLEMS

This is not the first time the data mining technique would be used to solve one problem or the other. The essence of it application this time around is to use its to classify the students results based on the available data in order to measure the accuracy of their classification.

2.1 Data Mining Process

There are different types of Data Mining process available according to [1] but in this paper, the cross industries standard process mining simply refer to as CRS-DM) shall be focused upon and it will be describe briefly as follow. It involved the following stages.

1. **Business objective:** With reference to [4] this deal with definition of the objective involves in carrying out mining activities i.e. these data to be involved.
2. **Data Understanding** According to Linard this is the second stage of data mining process and it involve the under study of data to be used in data mining process.

3. **Data Preparation:-** This is the third stage according to [5] and it involves clean of data remove wanted items from the data available.
4. **Modeling:** according to pans- Nins and Co 2005, it deals with development of model with any chosen programming languages.
5. **Testing of Data:** It involves testing of model that it is developed in above stage with data and see how perfect it.
6. **Evaluation:** according to & Co 2005 it deals with evaluation of the modeling process.
7. **Documentation:** (LN 2005) it involves the documentation of the previous stage and that is possible if the model that is developed is tested to be perfectly alright.

The diagrammatical representation of data mining process can be given below according to [7].

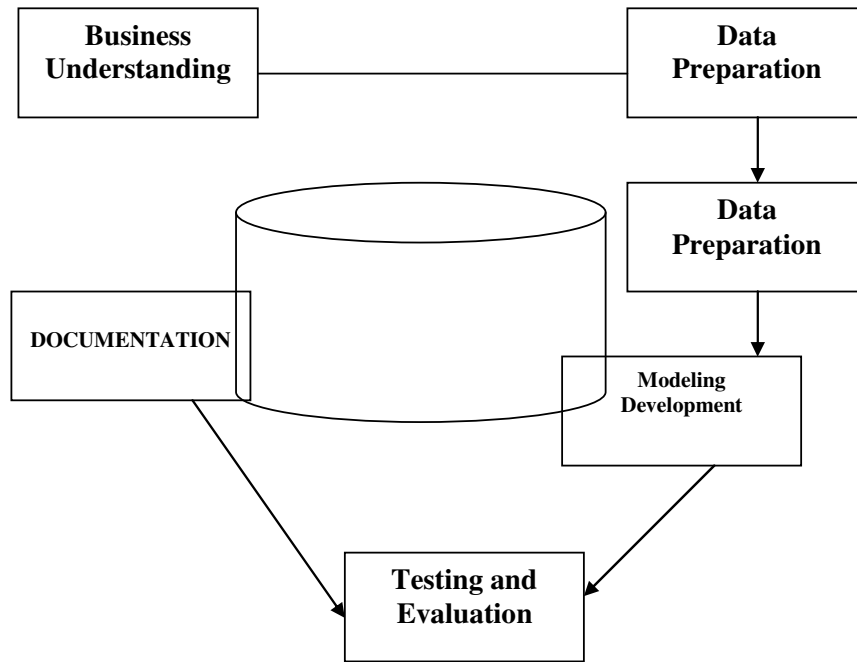


Figure 1.2: CRISP-DM process

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 based algorithm and it is given below the C program was developed to assist the indetermination of TPR, TFR, MCR and AR.

Based 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
3. Stop

The preceding algorithm can be interpreted in the confusion matrix as shown in the table below.

Table 1.1: Confusion Matrix

	PASS	FAIL
TRUE	TP	TF
FALSE	FP	FF

The above confusion matrix (2 x 2 matrix) according to Trevor & Robert (2009) can be used 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.

$$TPR = \frac{TP}{TP + FF} \times 100 = \dots\dots\dots \text{equation 1}$$

$$TFR = \frac{TF}{TF + FP} \times 100 = \dots\dots\dots \text{equation 2}$$

$$MCR = \frac{TP + TF}{TP + TF + FP + FF} \times 100 \dots\dots\dots \text{equation 3}$$

Where

Accurate classificcate rate =
 1 – MCR

Experiment 1

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

TR = 2
 FF = 16
 TF = 1
 FP = 6
 $TPR = \frac{TP}{TP + FF} \times 100 = \text{equation 2}$
 $TP = \frac{2}{2 + 16} \times 100$
 $= 2 \times 100 = 11.11\%$
 18

Experiment 2

The table 1.3 can also be used to calculate TPR, TFR, MCR and accuracy as follow;

TP = 3
 TF = 2
 FFD = 9
 FP = 13
 $TPR = \frac{TP}{TP + FF} \times 100$
 $= \frac{3}{3 + 9} \times 100 = 300$
 $= 25\%$

$TFR = \frac{TF}{TF + FP} \times 100$
 $= \frac{2}{2 + 13} \times 100$
 $= 200 = 13.3\%$
 15

$$\begin{aligned} \text{MCR} &= \text{FP} + \text{FF} \\ &= \text{FP} + \text{FF} + \text{TP} + \text{TF} \\ &= 13 + 9 \times 100 \\ &= 22 \times 100 \\ &= 27 \end{aligned}$$

$$\begin{aligned} \text{Accuracy} &= 100 - 81.5 \\ &= 18.5\% \end{aligned}$$

Table 1.2: NDIII Business Studies Result

KWARA STATE POLYTECHNIC, ILORIN
 CENTRE FOR CONTINUING EDUCATION
 DEPARTMENT OF BUSINESS ADMINISTRATION
 03/12/08 2006/2007 DIPLOMA/CERTIFICATE RESULT
 10:21 ND III BUSINESS STUDIES (PART TIME)

DIPLOMA/CERTIFICATE SUMMARY	RESULTS	GRADE	GRADE POINT	RANGE	C.G.P.A	CLASS
TOTAL NO. OF STUDENTS	= 95	A	4.00	75% 100%	3.50 – 4.00	DISTINCTION
TOTAL NO. WITH DISTINCTION	= 0	AB	3.50	70% 74%	3.00 – 3.49	UPPER CREDIT
TOTAL NO. WITH UPPER CREDIT	= 2	B	3.25	65% 69%	2.50 – 2.99	LOWER CREDIT
TOTAL NO. WITH LOWER CREDIT	= 76	BC	2.75	55% 59%	2.00 – 2.49	PASS
TOTAL NO. WITH PASS	= 16	C	2.50	50% 54%	Below 2.00	CAN'T GRADUATE
TOTAL NO. WITH C'OVER	= 1	CD	2.50	50% 54%		
TOTAL NO. TO REPEAT	= 0	D	2.25	45% 49%		
TOTAL NO. TO WITHDRAW	= 0	E	2.00	40% 44%		

Table 1.3 ND III Secretariat Studies Result

KWARA STATE POLYTECHNIC, ILORIN
 CENTRE FOR CONTINUING EDUCATION
 DEPARTMENT OF BUSINESS ADMINISTRATION
 16/12/08 2006/2007 DIPLOMA/CERTIFICATE RESULT
 15:08 ND III SECRETARIAT STUDIES (PART TIME)

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

4. DISCUSSION OF THE RESULTS

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

SOURCE CODE

```
#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; // FF for Part 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 |_____|\n";
cout << "\n |_____|MAIN MENU|_____|\n";
cout << "\n |_____|\n";
cout << "\n | 1. Get True Positive Rate | |\n";
cout << "\n | 2. Get True Failure Rate | |\n";
cout << "\n | 3. Get Marginal Misclassification Rate |\n";
cout << "\n | 4. EXIT | |\n";
cout << "\n |_____|\n";
cout << "\n | Select from Options (1 - 4) above: | |\n";
cout << "\n |_____|\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");
    printf("%s%f%s\n", "True Positive Rate for Full Time is ", tprf, "%");
    printf("%s%f%s\n", "True Positive Rate for Part Time is ", tprp, "%");

    cout << "\n\nGo Back to Mainmenu? Y/N:"; cin >> q;

    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");
printf("%s%f%s\n", "True Failure Rate for Full Time is ", tfrf, "%");
printf("%s%f%s\n", "True Failure Rate for Part Time is ", tfrp, "%");

cout << "\n\nGo Back to Mainmenu? Y/N: "; cin >> q;

if (q=='y' || q=='Y') {
    clrscr();
    mainmenu();
}

}

void mcrModule () {
    char q;
    q='N';
    clrscr();
    // Accept MCR Variables for Full Time
    printf("Enter the False Positive for Full Time: ");
    scanf("%f",&fpp);
    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",&fff);
    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 = (fpp * fff * 100) / (tpf + tff + fpp + fff);

    //Calculate MCR For Part Time
    mcrrp = (fpp * ffp * 100) / (tpp + tfp + fpp + ffp);

    //Display MCR result for Both Full Time and Part Time
    printf("\n\n");
    printf("%s%f%s\n", "Marginal Misclassification Rate for Full Time is ", mcrf, "%");
    printf("%s%f%s\n", "Marginal Misclassification Rate for Part Time is ", mcrrp, "%");

    cout << "\n\nGo Back to Mainmenu? Y/N: "; cin >> q;

    if (q=='y' || q=='Y') {
        clrscr();
        mainmenu();
    }

}
```

5. CONCLUSION, FINDING AND RECOMMENDATION

This Algorithm assists in determining the following:

1. The level of performance of the student can easily be measured
2. The level of the instructors performance can also be determined
3. The nature of examination can easily be determined.

6. FURTHER RESEARCH WORK

The reader is advice to work on another areas where rates of performance can be determined.

REFERENCES

1. David, L.O (2005): Advance Data MINING Technique, Spirnger-verkag, Berlin.
2. IAN witten and Eibe Frank (2000) Practical machine learning tools and Techniques with Java implementations published by McGraw Hill India.
3. Margaret .H. Dunham (2003) Data Mining Introduction and Advanced Topic, published by Derson Education, Inc India.
4. Ln, C.C & Siegel L.A (1988): Mathematics Applied to Deterministics problems in the natural sciences, India.
5. Peter Cabena, Pablo Hadjinian, Rolf Stadler, Japp verless Allesandro Zanari (1997) Data mining: from concept to implementation published by Johnwilly & Co.
6. Richard O. Duda, Peter E. Hart, David G. Stork (1980) pattern Classification published McGraw Hill New York.
7. Trevor Hastie, Robert Tibshirani and Jerome Fiedman (2001): The elements of Statistical learning published by John willy New York.
8. Yike Guo and Robert Grossman (1992) scaling Algorithm. McGraw Hill – New York.
9. Zhigang Xiang, TOROY A plastock and D.S Avaphand (1996): Computer graphics, published by TATA-Mc-Graw-Hill, New Delhi.

ABOUT THE AUTHOR



Olagunju Mukaila is a senior lecturer and head, Computer Science Department, Kwara State Polytechnic, Ilorin, Ilorin Nigeria. Olagunju obtained his first degree in Computer Science and Mathematics (Combined Honour), from Federal University of Technology Minna, Nigeria. He obtained his second degree in Mathematics with Computer Science option from University of Ilorin, Ilorin Nigeria. He also obtained his PHD in Computer Science from University of Ilorin, Ilorin Nigeria. Olagunju is a Computer and Mathematics expert, his area of specialization includes, Algorithm development and fluid flow pattern recognition. He has published Articles both in the National and International Journals. One of his popular articles is determination of time of death predictive model. Olagunju is a member of Computer Professional Council of Nigeria (CPN), member of Teacher's Registration Council of Nigeria and Nigeria Computer Society (NCS).