Predict Academic Performance of Students using an K Nearest Neighbour Algorithm Case Study: MATLAB Course

Noha Hassan Osman Ragab[1], Dr Saif Eldin Fattoh[2]

[1]Lecture, Faculty of Education of Afif, Shaqra University, Kingdom of Saudi Arabia

[2]Dean of Emirates Unversity, Khartoum, Sudan

Abstract: *The aim of the sudy predictability of varying female third level. In the academic year 1437department of chemistry at MATLAB subject and knowledge of the academic performance of female students in support of the educational process, use the search applied study in the area of detecting knowledge and techniques and data Mining their appropriateness and taking advantage of the volume of data, used Classification techniques represented in K-nearest neighbor (KNN) algorithms to take advantage of the grades for which obtained by female students in MATLAB subject in the previous two years (1435/1436) with the Grade Point Average (GPA) female students had been applied through WEKA tool that supports many of the latest technologies and algorithms in data Mining, and reached the search of many of the results, the most important of which is that the value of (K) in the algorithm based on the value of the Root Mean Square Error (RMSE), which lead to the predictability of the largest possible number of female students to be know the academic performance For them, The predictability of a course will help to assess the performance of members of the teaching staff and knowledge of the failures and negatives and work to address before the end of the semester.*

Keywords: data mining and discovery of knowledge, predictability, K-Nearest neighbour algorithm

1. Introduction

We find that many universities and higher institutes and all the institutions of education and environment is very rich in information and data which could be used in the development of the educational process by conducting scientific studies and research using the huge amount of data within those institutions as well as to accentuate those institutions flexibility which can be, a researcher from the performance of the discussed in a scientific way true. With the emergence of the flag of data, exploration has become possible to take advantage of the amount of data in the databases of universities to develop the educational process

1.1 Data Mining

The development of science and the economy and information technology and communications has led to an increase in the amount of digital data recently tremendous amounts of data is no longer the traditional analysis of statistical (for example) are able to deal with it. But this afternoon for prospecting and exploration for data (data mining) since the late 1980s and proved its presence as one of the successful solutions to the analysis of huge quantities of data, transferred just information accumulated and understood (data) to the valuable information that can be exploited and take advantage of (know).where he arose (Data mining) contribution of several specialties including statistics and databases and artificial intelligence. The exploration of data (sometimes called the discovery of knowledge) is the process of analysing data from different perspectives and draw links and useful information, such as the information that could contribute to an increase in profit, reduce costs, or both.Or is the process of disclosure and find information of interest through the use of a range of complex tools. Some of these tools include statistical tools habitual

residence and artificial intelligence and charts of the decision-making computer.[3] through the above could say that (Data mining) based on the following:
- Data are the facts and figures and the texts which could be dealt with by the computer.
- Information: models and relations between those data which constitute useful information.
- Knowledge: previous information can be diverted to know about the historical patterns or prospects for the future. [5].

1.2 K-nearest neighbor algorithm

Is prospecting techniques aimed at predictability through compared records similar to the Register to be predictability to the estimation of the Unknown value for the Register, on the basis of information to those records[4][2]. In this study was the use of this algorithm to predict the success or failure of female students in a course is done using the information for course and the students, the identification of a number of pilot records and then use the aim of predictability required value.

Steps for K nearest neighbour algorithm[5][7]
- Select the number of the **nearest** neighbors let K.
- Calculate Euclidean distance between the Register to be predict the value and the nearest neighbor.

$$d(x_i, x_k) = \sqrt{\sum_{t-1}^{n} [x_{ti} - x_{tk}]^2}$$

- Distance arranged to give its level of the smallest distance to the highest walking, then select the closest neighbours based on the distance to a minimum. K-th.
- A whole category closest neighbours.

- The calculation of mean of nearest neighbours forecast of the situation of the Register to be unpredictable.
- Continue assessing a function of the objective of the records to be unpredictable.
- Calculate RMSE value (Root Mean Square Error) for each of the value of the K if less than the past stop if not go to Step 1.

2. The Problem of Research

There are problems and difficulties faced some educational institutions in evaluating the performance of academic student is the difficulty of access to a comprehensive evaluation of the results to be rectified may be moral and Abstract general rather than felt and quality, it may be difficult to determine the results of the learning that we want to be measured in clear and precise, are still many tools of measurement and evaluation in some universities inaccurate, which leads to the weakness in the interpretation of the results and analysis. The assessment of student performance Academic and aims to provide the capacity to exploration and to focus on the most important information in the register of student and help in the future forecasts student allowing the estimation of the correct resolutions and at the appropriate time, contribute to the development of criteria for the academic performance of the student and members of the teaching staff of the university in general.and therefore the problem this to predict the final degrees to take the right decisions.

3. The Importance of Research

The importance of research to benefit in:
- Technical use new to measure the academic performance of the student at the university.
- The analysis of the data to reach new results help in the development of the educational process.
- Assistance in improving the student performance the Accademia Gallery.
- The use of the results of the study in the development of the academic performance of a member of the Al-courses.

4. The Objectives of the Research

Research tries to achieve the following objectives:
- To know whether the best choice for the algorithm claims to the predictability of varying degrees of female students.
- To know the role of predictability of the academic performance of female students in the curriculum to help improve the method of teaching and education outputs.
- The extent of the contribution of the predictability of the decision of the curriculum to assess the performance of members of the teaching staff.

5. The Study Questions

- Is the best choice of the gradient of algorithms and lead to the predictability of varying degrees of female students in minutes?
- What is the role of the predictability of the academic performance of female students in the development and improvement of the educational process?

- Do you predict the decisions and assist in the evaluation of members of the teaching staff?

6. The Research Methodology

In this study was the use of the descriptive embracing analytical and statistical applied where the development of a method of assessing the academic performance of the student on the list of the methods currently used in the drilling systems data (data mining Systems).This method aims to build a decision of the forecasts of relevant to the assessment of student Accademia Gallery this reduces the problems resulting from the weakness of the methods used in the arrangement of the decisions it has been using the algorithm closest neighbour (K-Nearest Neighbour) were applied to the degrees of female students.

7. The Limits of Study

Substantive session: the application of techniques of prospecting data on final degrees of a female of the Chemistry Department of the Faculty of Education of Afif in MATLAB Subject for the years (1435 &1436) as well as to the test of the middle and the quarterly ratio and attend the lectures. The spatial resolution: Faculty of Education of Afif of the University of Shaqraa' in the Kingdom of Saudi Arabia. Temporal: the first semester of the year 1437.

8. Theoretical framework:

MATLAB subject Special Rapporteur third level within the plan of the department of chemistry at the Faculty of Education Of Afif - Shaqra University - the Kingdom of Saudi Arabia.

It has scheduled a variety of units that may help the student to solve the problems facing the development of the knowledge capacity they have.

The distribution of degrees decision depending on academic appropriation in the Kingdom of Saudi Arabia and shall be as follows:

test first midterm 15% and second midterm 15% attend lectures and 10%practical testing 20% final testing 40%.

9. Related Work

We will refer to some of the studies linked to the search directly or indirectly where many of the studies on the level of the world the application of algorithms for data mining to discover knowledge in universities, in the study [1] the application using the data mining on the students of the University of Science and technology to detect patterns which help decision makers in improving the educational process.

The study interested in [5] applications of data mining in the field of higher education, focusing on inputs of the educational process and outputs and how the impact of each other, a method is used neural networks for data mining, the results show a variety of relations between the curricula,

watches, adopted the nature of students and graduates and the functions that they occupy, in addition to the conclusions of other useful decision maker in universities

Studies in [6] [7] been foreseen university performance for students on the basis of personal data for students and use data mining algorithms from the algorithm k nearest neighbor. Most of the previous studies used data mining algorithms to predict the academic performance of students in general through scientific career in the university the expectation successfully students or students in a curriculum is not used one algorithm but several and algorithms compared between them in the recent results of those studies to reach the best predictable algorithm the academic performance of the students in a Subject curriculum it is k nearest neighbour algorithm

10. The Methodology

Collection data of female of the Chemistry Department of the Faculty of Education of Afif of attendance MATLAB subject , the first midterm , second midterm and Grade point Average(GPA) for the years 1435 -1436 the work of a model to be by predict the academic performance of female students in the year 1437 the first semester and enter data on the program (Microsoft Excel 2010) and then data saved on the type of files (CSV) to be entered into the weka program and the establishment of a model of predictability for use in forecasting the academic performance of female students at the MATLAB subject for the first semester 1437.

10.1 A Society Study

Female students of the Chemistry Department of the Faculty of Education of Afif - Shaqra university-king of Sudia Arabia.

10.2 Study Sample

Female third level for the first semester of the year 1437 department of chemistry.

10.3 Study Tool

10.3.1 Model building and data collection.
The data were collected to build the model in the program Excel 2010 and then saved gender .csv file the data is the degrees of female students in MATLAB subject for the previous two years 1435 and1434 contained requesting name and number on the university campus, med1, med2.attendence and GPA.

Table 1: Data Model

Field name	Data type	Value
Student name	Text	Text
Student number	Number	Number
Middtrem 1	Number	(1-20)
Middterm 2	Number	(1-20)
Attends	Number	(0-5)
GPA	Number	(2.5-5)
Final Exam	Yes, No	Yes, No

10.3.2 Model Building used weka Program K-Nearest Neighbor algorithm known to Lazy, IBK in Weka program were extracted the following form:
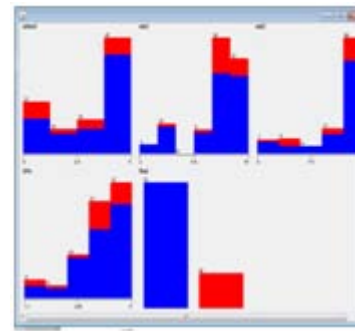
11. The Results and Analysis

First
when designing the model was choose the model (K=3) on the basis of the Root Mean Square Error (RMSE) because it gives the highest value in the expectation as indicated in the table below

Table 2: RMSE value

k	Root Mean Squered Error(RMSE)
1	0.4274
2	0.4192
3	0.4306

On the basis of the Model predicted the final degrees of female students as follows:-

Table 3: Number of predicted

K=1		K=2		K=3	
Fail	Pass	Fail	Pass	Fail	Pass
10	16	8	13	9	19

Graphic shows the following day that the choice of the value (K) affect the number of female students to predict their success or failure in the decision of the MATLAB Courses.
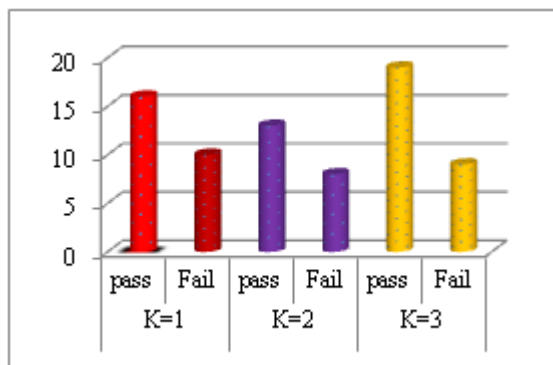
12. Conclusions

In this study was to identify how to use KNN algorithm closest neighbor in predicting the success or failure of female students in the A course, the study reached the following conclusions:

1) The best choose for the value of the (k) In the algorithm leads to predicting more than a possible number of female students to know the academic performance of them. Find at our choice of value,(k=1) Total number of women students is 28 female students were forecast for 26 students only. When the value (K=2) predicted the number 21 requesting only when the value (K=3) predicted for all female students.

2) The knowledge of the academic performance of female students in a curriculum help members of the teaching staff to know the points of weakness and strength of female students, it also helps to improve the method of teaching and education outputs.

3) The predictability of a course will help to assess the performance of members of the teaching staff and knowledge of the failures and negatives and work to address before the end of the semester.

13. Recommendations

- Must be put to predict the academic performance of students the basic measure to improve the operation and development of instruction.
- Making the predictability of the academic performance of students in courses is the basic criterion to upgrade for the education and learning.
- the opening of the researchers in data mining techniques and data to take advantage of this modern science and the improvement and development of educational institutions and outputs.
- The universities and institutes and educational institutions have enormous stores data must take advantage of this data to discover the patterns and the knowledge that help to improve education for the educational process.

References:

[1] Ahmad Mohammad Sheikh Ali Alatas warehouses distributed data to assist in decision-making - Application of the University of Science and Technology is a "master of allocated Computer Sciences-2005M.

[2] Obeid, Mustafa(2005),"prospecting in databases and explore the hidden information, „‟center for research and studies and training‟‟.

[3] Jiawei han & Micheline Kawber, "Data Mining Concepts and Techniques", third Edition, University of Illinois at Urbana, 2003.

[4] Chakraborty, Dipanjan(2008), K-Nearest Neighbor Learning .

[5] Z. Pardos, N. Hefferman, B. Anderson, and C. Hefferman, "The effect of Model Granularity on Student Performance Prediction Using Bayesian Networks," Proceedings of the international Conference on User Modelling, Springer, Berlin, pp. 435-439, 2007.

[6] D. Kabakchieva , K. Stefanova, V. Kisimov, Analyzing University Data for Determining Student Profiles and Predicting Performance, Proceedings of the 4th International Conference on Educational Data Mining, Eindhoven, the Netherlands. July 6-8, 2011.

[7] D. Kabakchieva, Predicting Student Performance by Using Data Mining Methods for Classification, Cybernetics and Information Technologies, vol. 13, No. 1, pp. 61-72, 2013.

[8] Ki-Yeol Kim,Byoung-Jin Kim and Gwan-Su Yi(2004), Reuse of imputed data in microarray analysis increases imputationefficiency, BMC ioinformatics,Vol.5.