Prediction of Students' Admission Rates Using SMOreg Algorithm

Shazali Siddig Mohammed
Faculty of Computer Science
Information Technology Department
AL-Neelain University
Khartoum – Sudan
shazali33@gmail.com

Ali Abdallah Abaker
Faculty of Computer Science
Accounting Information Systems Department
AL-Neelain University
Khartoum – Sudan
misterali88@gmail.com

***Abstract* -The purpose of this paper is to present the current and future trends for computer science disciplines at public and private Sudanese universities using admission data from higher education, and demonstrating the most preferred fields of study in computer science and their future trends. SMOreg algorithm in Weka mining tool was used to predict the number of students in the field of computer science in Sudan for two years ahead 2018, 2019, and illustrating the students' distributions in computer science departments.**

***Keywords* -** Predictive Analysis, SMOreg algorithm, Educational Data Mining

1. INTRODUCTION

Data Mining means discovering useful information from large amount of data[1] , also known as Knowledge Discovery from Databases, data mining has become one of the most interesting research area for researchers as data is increasing rapidly every day in database systems and more systems have been computerized data mining looks for how can we make use of this data, and many mining techniques and algorithms have been developed [2].

Educational Data Mining (EDM) is an emerging field of data mining aims to investigate the educational system. EDM has become an independent field of research [3], 40 tools in the field of EDM have been counted [4]. As the number of students that join higher education in Sudan is increasing very fast, in addition to the E-learning tools and techniques have opened new doors to gather students' data for data mining research [5].

The importance of this analysis is to present the trends in the computer science disciplines, and predicting the number of students for the academic years 2018, 2019. Also this analysis will show the distribution of students with their departments, this will help the ministry of higher education to plan for resources distribution, determines the number of students to be accepted in a particular department, in addition to make future forecast about graduates for employment.

2. EXISTING LITERATURE

P. Thakar et al, conducted a survey paper [6]. And they present a comprehensive review on educational data mining from (2002-2014) and classified EDM into five areas:

1. Survey of published paper in educational data mining.
2. Predicting student's academic performance.
3. Comparison of Data Mining Techniques.
4. Correlation among factors.
5. Other areas of Education.

Weka is the software used for data Mining Found only on the islands of New Zealand and it is open source software issued under the GNU General Public License. The version is 3.9.2 and the forecast package was downloaded and installed.

3. DATA SOURCE

This dataset was collected from the Sudanese Ministry of Higher Education and Scientific Research (MHESR) – Admission Administration. The Dataset includes Students admission for public and private Sudanese Universities for three years (2014-2015, 2015-2016,2016-2017), it contains the following attributes (Admission year, Seat No, Admission Type ,Department, College, University). The dataset has 349,045 instances of Admitted Students in Bachelor Degree programs.

4. DATA PREPROCESSING

Computer science students 15770 have been extracted from the data, because this paper addresses the field of computer science. About 100 instances have been removed from the dataset and remains15670 instances, seat No, College and admission type have been removed**.** A new attribute (university state) was added.

5. METHODOLOGY

In this paper Sequential Minimum Optimization Regression (SMOreg) was used as a prediction algorithm for time series Data(admitted students) for Sudanese universities for computer science faculties.[7][8]have achieved the best results using SMOreg algorithm.

6. EXPERIMENTS

The Data set was uploaded into Weka Mining software, sequential minimal optimization regression **(**SMOreg) forecasting algorithm was executed for two years ahead, the results show that the predicted number of students for 2018 will be 9,329 and for 2019 will be 13,869 as it is illustrated below in the chart with sequential increase over time. For the evaluation Mean Absolute Error is 2.76 and the Root Mean Squared Error is 2.76.

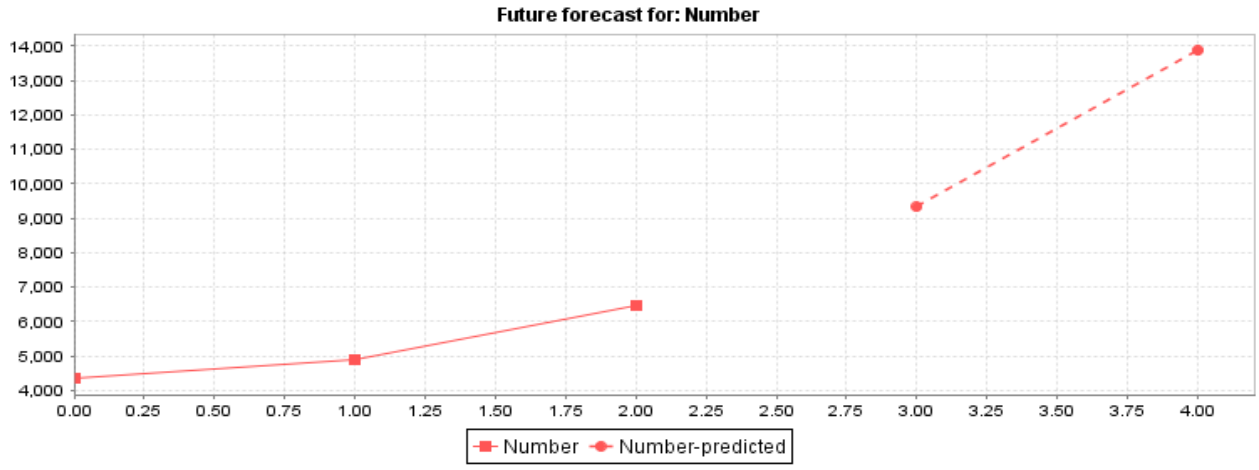

Fig 1. Shows prediction of Admitted Students to Computer Science Faculties for 2018, 2019.

Fig1. presents the predicted number of students for two years ahead as the chart illustrates the increasing number of students, executed by Weka software and SMOreg algorithm for producing the chat above.

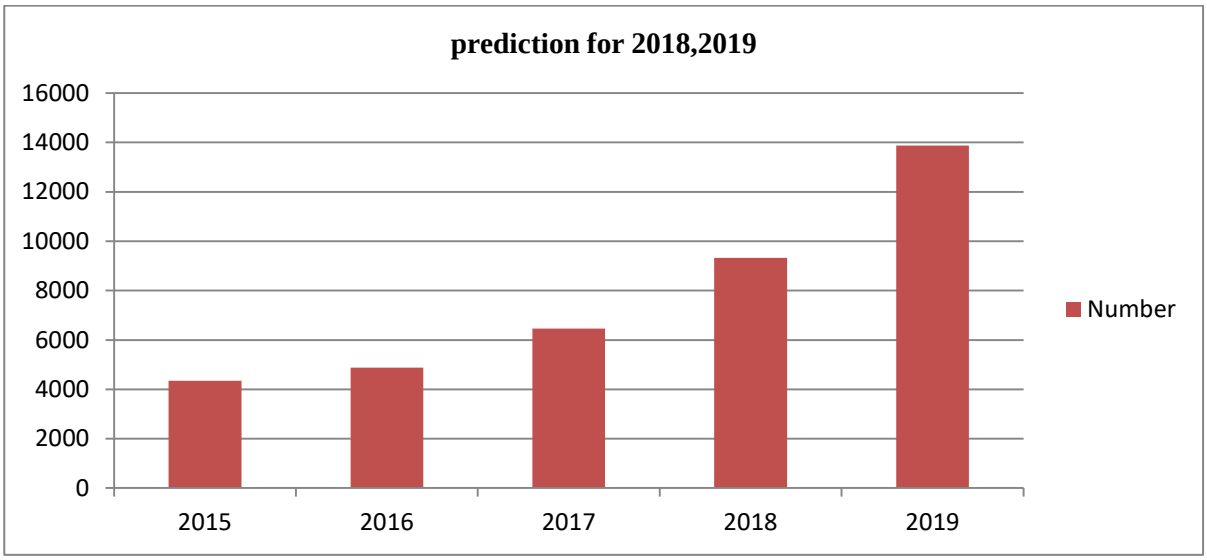

Fig 2**:** Shows prediction of Admitted Students to Computer Science Faculties for 2018, 2019.

Figure 2. above presents the historical and the predicted number of students at public and private Sudanese Computer Science faculties, the admission is increasing significantly every year currently it is 6454 in 2017, and it's expected to double by year 2019 to be approximately 13869.

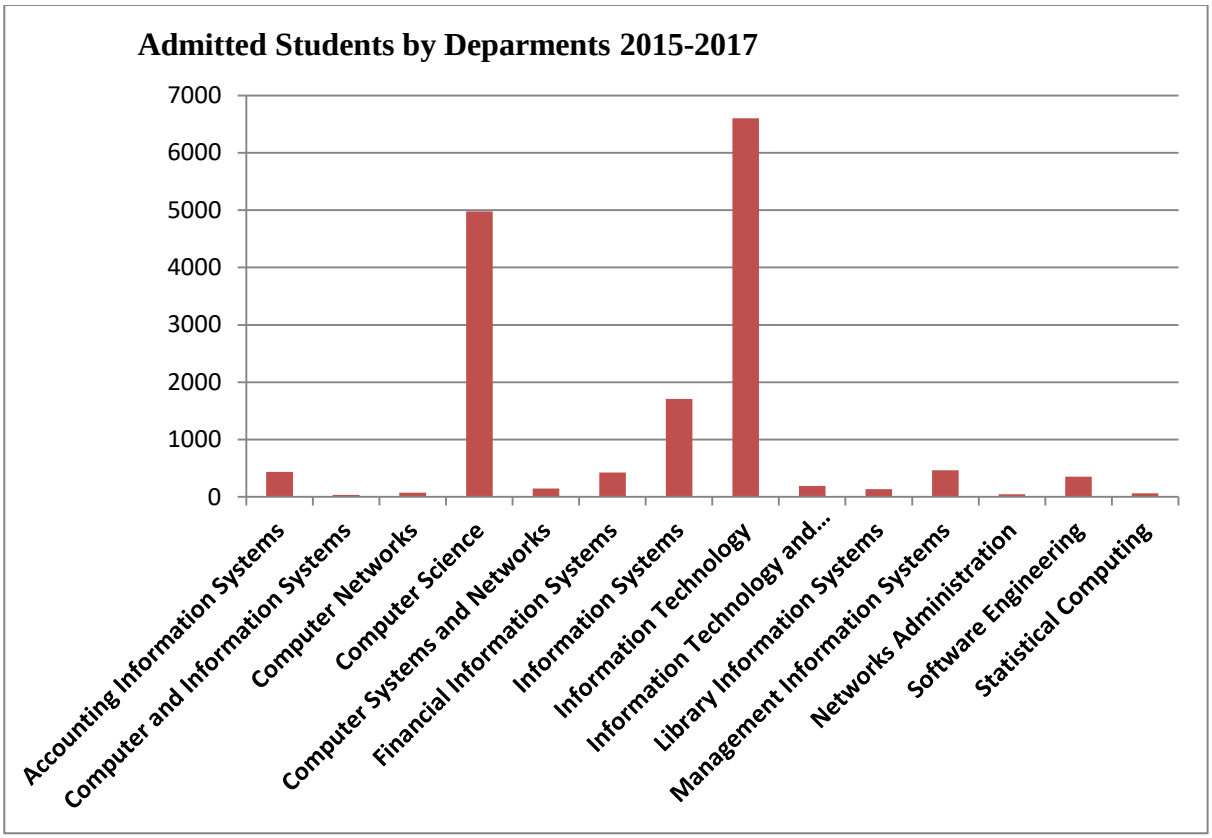

Fig 3 Shows Departments and Students numbers from 2015-2017.

From the figure above it is clear that Information technology Department is the most preferred followed by Computer Science and Information Systems for the last three years in Sudan. But departments like Accounting Information Systems, Management Information Systems, library Information Systems and Software Engineering with less students and are new growing fields of study.

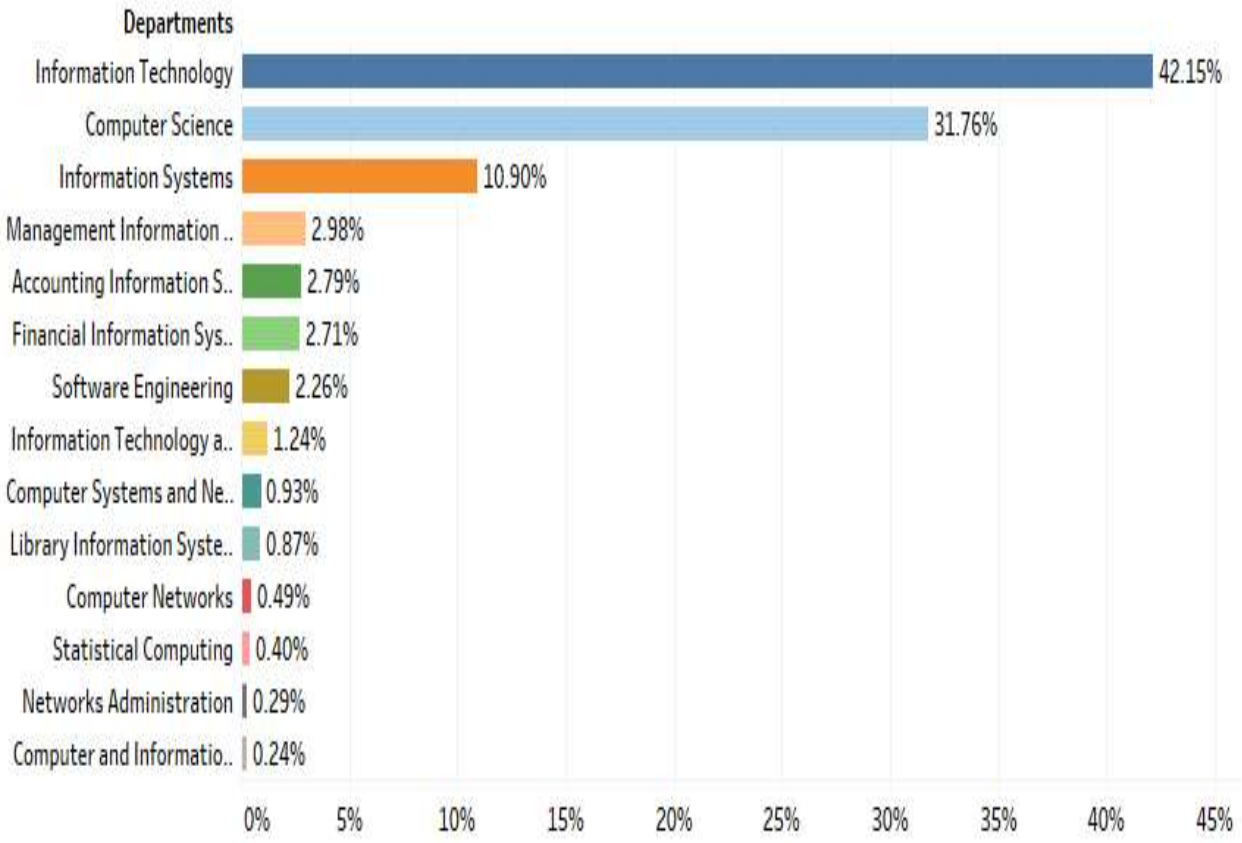

Fig 4 Shows departments and admitted students rates classified by departments.

From the figure above its clear that Information technology Department has 42.15% of all students who are studying in computer science disciplines followed by Computer Science with 31.76% and third Information Systems 10.90% for the last three years in Sudan. And Accounting Information Systems, Management Information Systems, library Information Systems have about 2% and Software Engineering with 2.26%.

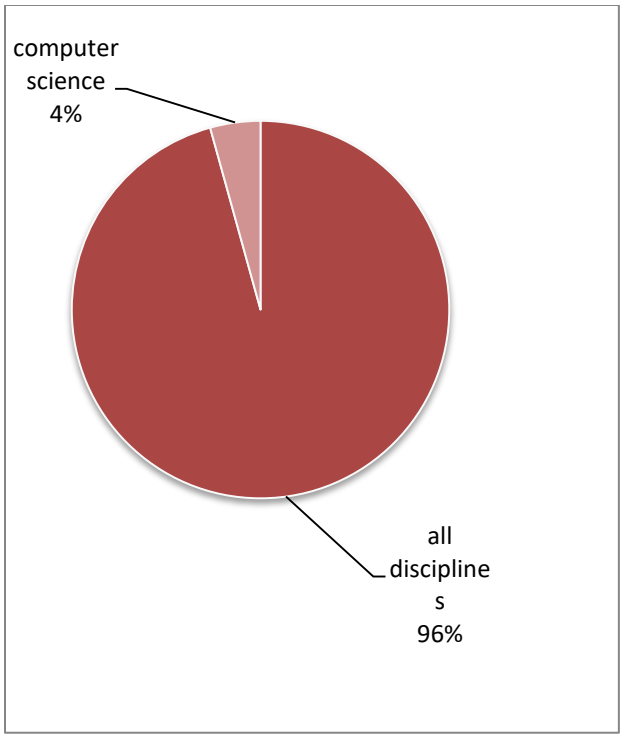

Fig 5 illustrates computer disciplines and all fields.

From the figure above computer science students are representing about 4% comparing to all students in the other disciplines, and this percent is expected to increase in the coming years as shown in Fig 2.

7. RESULTS

- It's expected that the number of students in computer science disciplines will approximately be about 13,869 by 2019**.**
- It's clear that from figure 5 information technology department is the most preferred with 42.15% of total students in computer science disciplines.
- From figure 4 computer science is the second with 31.76%.
- From figure 3 there are 14 distinct departments relevant to computer science disciplines in Sudan.
- Students in computer science disciplines are representing 4% of the total admitted students for higher education in Sudan.

8. FUTURE WORK

- As the number of students is increasing more data will be generated ranging from admission data, student behavior, students' interactions with

educational systems and online learning, results, and graduation all these types of data need to be collected and analyzed.

- Higher Education needs investment so Educational Data Mining will provide accurate information for decision making.
- Educational systems are developing and new academic methods are being applied so the future role of data mining is to measure the impact of these methods.
- Such study can be conducted in other disciplines like medicine, engineering.

9. CONCLUSION

To conclude predictive analysis will become a strong tool in decision-making, these future predictions results can Support the Ministry of Higher Education for planning as it is expected to receive more students in computer science disciplines. The authors wish this paper to draw researchers' attentions on the promising field of educational data mining in Sudan.

ACKNOWLEDGEMENTS

Our thanks to the Ministry of Higher Education and Scientific Research especially the Admission Administration for providing us with the data set which was vital in completing this research.

REFERENCES

[1] G. B. Tarekegn and V. Sreenivasarao, "Application of Data Mining Techniques to Predict Students Placement in to Departments," *Int. J. Res. Stud. Comput. Sci. Eng.*, vol. 3, no. 2, pp. 10–14, 2016.

[2] S. Ventura and C. Romero, "Data mining in education," vol. 7, no. October 2015, pp. 43–48, 2012.

[3] R. S. J. D. Baker, "Data mining for education," *Int. Encycl. Educ.*, vol. 7, pp. 112–118, 2010.

[4] S. Slater, S. Joksimovic, V. Kovanovic, R. S. Baker, and D. Gasevic, "Tools for Educational Data Mining: A Review," *J. Educ. Behav. Stat.*, vol. XX, no. X, pp. 1–22, 2016.

[5] S. L. Prabha and A. R. M. Shanavas, "Educational Data Mining Applications," *Oper. Res. Appl. An Int. J.*, vol. 1, no. 1, pp. 23–29, 2014.

[6] P. Thakar, A. Mehta, and Manisha, "Performance analysis and prediction in educational data mining: A research travelogue," *Int. J. Comput. Appl.*, vol.

110, no. 15, pp. 60–68, 2015.

[7] V. Choubey, “Time Series Data Mining in Real Time Surface Runoff Forecasting through Support Vector Machine,” vol. 98, no. 3, pp. 23–28, 2014.

[8] Saigal S and Mehrotra D, “PERFORMANCE COMPARISON OF TIME SERIES DATA USING PREDICTIVE DATA MINING TECHNIQUES Advances in Information Mining,” vol. 90939093, no. 1, pp. 57–66, 2012.