World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:11, No:11, 2017

Development of the Academic Model to Predict Student Success at VUT-FSASEC Using Decision Trees

Langa Hendrick Musawenkosi, Twala Bhekisipho

***Abstract*—**The success or failure of students is a concern for every academic institution, college, university, governments and students themselves. Several approaches have been researched to address this concern. In this paper, a view is held that when a student enters a university or college or an academic institution, he or she enters an academic environment. The academic environment is unique concept used to develop the solution for making predictions effectively. This paper presents a model to determine the propensity of a student to succeed or fail in the French South African Schneider Electric Education Center (FSASEC) at the Vaal University of Technology (VUT). The Decision Tree algorithm is used to implement the model at FSASEC.

***Keywords*—**Academic environment model, decision trees, FSASEC, K-nearest neighbor, machine learning, popularity index, support vector machine.

I. Introduction

THE success rate in academic environments is not only a concern for those institutions but also governments, sponsors such as the public and the private sectors, parents, students themselves and other stakeholders. It is therefore fitting to investigate the propensity of those students to succeed using scientific methods such as machine learning. Machine Learning (ML) has a variety of algorithms that can be applied in addressing this problem. The South African government and funders can save a lot of resources when funding these institutions. Therefore, the application of rigorous methods of ML can improve the efficiency in the academic sector.

For the most part in South Africa, the largest contributor of funding in public education is government, that is, the ministry of education. Although the Ministry of Education takes no account of income that is raised from student fees and other private sources, these public institutions have to account by submitting annual financial statements which reflect all income and all expenditure from all public and private sources [5].

The need to attract and retain students in engineering programs [8] remains by necessity, a focal point of interest and effort in engineering education. All universities and colleges have marketing departments to make sure that they attract the best of the best. They run various marketing programs for this purpose. Paul and Cowe Falls [2] highlighted the three aspects for engineering careers success based on the availability of the resources. Firstly, lifelong learning is fundamental for success in the 21st century engineering career. Staying abreast with the most recent technological advancement is essential for being innovative and creative. Secondly, a study in the engineering construction industry is the most critical aspect of fostering a successful career path was in developing a career network. This includes networking, mentorship training and constructive feedback. Thirdly, the aspect of engineering career success relates to the models "proactive personality" variable.

Langa Hendrick Musawenkosi is with the Department of Electrical Engineering, Vaal University of Technology, Vanderbijlpark, South Africa (e-mail: hendrickl@vut.ac.za).

Twala Bhekisipho is with the Department of Electrical & Electronic Engineering Science, University of Johannesburg, Johannesburg, South Africa (e-mail: btwala@uj.ac.za).

II. Overview of the Decision Tree Algorithm

A. Decision Trees (DTs)

DTs are simple yet successful techniques for supervised classification learning [7]. This classification method consists of decision nodes, connected by branches, extending from the root node until the terminating leaf nodes [1]. Starting at the root node attributes are tested at the decision node, with each possible outcome resulting in a branch.

The decision tree is essentially a structure that will split data points or categorize data points into different decisions. There is a question in each node in order to make a decision. The first step is to understand whether data is numerical or categorical. For example, university admission prediction service has been a challenging decision process of helping the right students to enter the right universities. This evaluation process in the past was attempted by linear programming models, regression formulas and neural networks [9].

In classification, there are many different methods and algorithms possible to use for building a classifier model [3]. Some of the popular ones would the k-nearest neighbor (kNN), artificial neural network (ANN), support vector machines (SVM) and logistics regression methods.

III. Other ML Algorithms

A. The K-Nearest Neighbor Classifier (kNN)

kNN is an algorithm that falls under the category of supervised learning algorithms. The classification as per this algorithm is done based on the distances between the training data and the testing data [6]. kNN is an example of instance based learning, in which the data set is stored, so that the classification for a new unclassified record may be found by

simply comparing it to the most similar records in the training set [1].

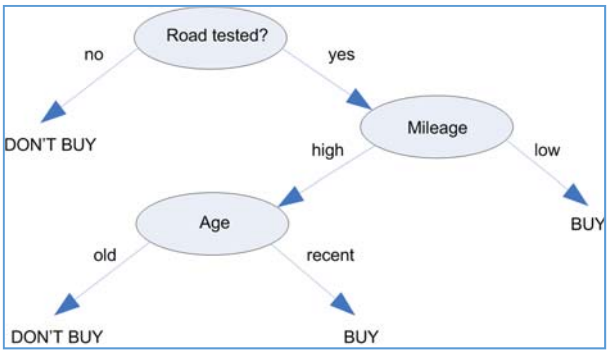

Fig. 1 Simple decision tree for buying a car

The kNN algorithm [11] is the earliest researched algorithm used for classification and is proved as one of the algorithms which have good classification results in Reuter data sets (including 21,450 and Apte data sets), but there are still some problems that need to be attended to. For example, it is not yet settled, how to select the value of k and how to select feature sets to make better the classification and their impact on each other.

B. The Support Vector Machine

SVM is an algorithm that uses nonlinear mapping to transform the original data into a higher dimension, [4]. SVM's are pattern classifiers [10] that can be expressed in the form of hyperplanes to discriminate between positive instances and negative instances pioneered by Vapkin.

IV. Modeling an Academic Environment

A. The Academic Environment Model

The environment is the sum total of surroundings of a living organism including natural forces and other things, which provide conditions for development and growth as well as danger and damage. An academic environment where a student exists; in order to model this, it is necessary to gather information about the student, the lecturer and the module which will form part of the environment.

It is clear from the diagram above that the academic or learning environment is composed of three parts, namely, the student, the lecturer and the subject. Each of these components has an impact on the outcome of the academic performance of the student.

The model for making the prediction represents the ML algorithm and can be written as:

$$F(x_{1:i}, y_{1:j}, z_{1:k}) = class(c_1 c_2, ...c_n) \quad (1)$$

where: $x_{1:i}$ = list of lecturer attributes, $y_{1:j}$ = list of the subject attributes, $z_{1:k}$ = list of the student attributes, c_1, c_2, …,c_n = class.

Obviously, some models will be more appropriate than others and some models will be more accurate than others. The decision tree has been selected in this paper.

Fig. 2 Academic Environment Model

B. The Lecturer's Popularity Index

One of the common problems in higher education is the evaluation of the instructor's performances in a course. The popularity index has been developed in this research to measure the performance of the lecturer or subject. Students that are taught by a specific lecturer have the opportunity to actually evaluate the lecturer personally. The percentage score of likes for a given the subject or the lecturer is given by the following equation:

$$L(x) = \frac{\sum_{i=1}^{n} x_i}{nk} \quad (2)$$

where: n = number of instances of likes for a lecturer or subject and k = the highest number of likes in an instance.

Clearly, some lecturers are more popular than others; there are lecturers whom students really detest and there are lecturers whom they adore. These can be due to several reasons, such as the appearance, teaching style, level of education, leadership and so on. The popularity of the lecturer has a correlation with the performance of the student. Similarly, the popularity of the subject can be measured.

C. The Academic Environment Client System

The panel below allows the student to rate the lecturer using choices of numbers between 1 and 5. If the lecturer is least popular then the choice would be a 1 and if the lecturer is a student's favorite then the choice would be a 5. This information is then captured in a database for future references. The popularity of the lecturer can increase or decrease with time depending on the performance of the lecturer. This is an import feature to have in the design.

A survey of 24 students was completed, where seven FSASEC staff members and four subjects were evaluated using the system as shown in the Fig. 3. Upon analyzing results, interesting observations were noted.

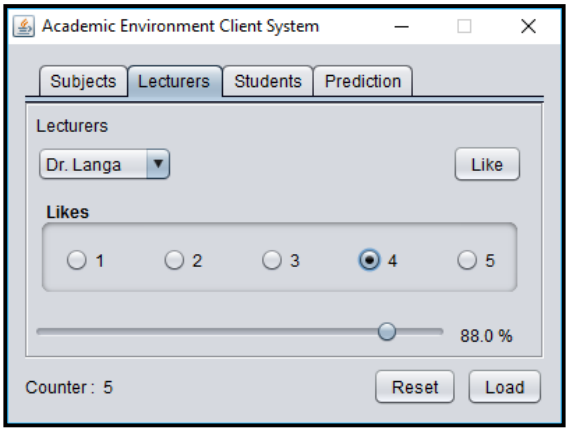

Fig. 3 Academic Environment Client System

D. Popularity Indices for FSASEC Lecturers

It is clear from the graph that some lecturers are more popular than others and that the highest popularity index is 82.90%, whereas the lowest popularity index is 46.40%. The highest index suggests that the students are particularly fond of the staff member, while the lowest popularity index suggests that we need to be concerned about the students' unfavorable reaction in this case. Nevertheless, we have a good measure of the perception of students about their lecturers and this is a fair enough gauge in terms of the quality of lecturers employed in the department of FSASEC.

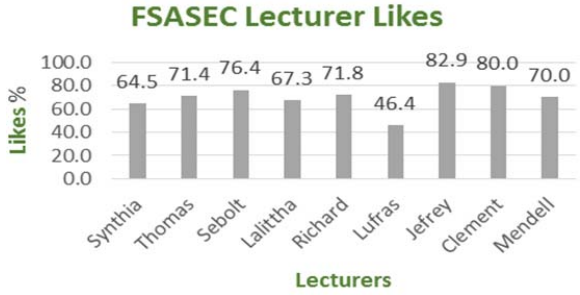

Fig. 4 Popularity Indices for FSASEC Lecturers

There are various reasons why a lecturer would be unpopular. It could just be shear laziness on his part, lack of understanding of the subject he teaches, their attitude, being too strict, and so on, to name a few but a few. And there are various reasons why a lecturer could be considered as popular. It could be that they are good in the subject matter, they have good qualifications, they are lenient, their attitude, again, to name a few. Evidently, some lecturers are more popular than others in this department as would be the case with other departments.

E. Popularity Indices for FSASEC Subjects

It is clear from the graph that some subjects are more popular than others.

The lowest popularity index has been for English Communication, 47.3% and the highest has been for Mathematics, 81.8%. Again, there are various reasons why a subject would be unpopular while another is more popular. If the subject is popular, it could be because it is considered easy, it could be well understood, etc., and if the subject is disliked, it could due to the fact that it is difficult to understand, the lecturer is not good at teaching it, or it is too abstract for students to understand and so on.

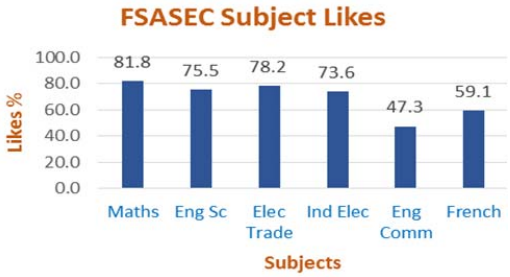

Fig. 5 Popularity Indices for FSASEC Subjects

V. Academic Environment Using the Decision Tree

The model is developed on the basis. Some algorithms will be more accurate than others and some will be more appropriate than others also.

TABLE I
ACADEMIC ENVIRONMENT MODEL FOR FSASEC – VUT DTS

Lecturer	Lecturer Likes	Subject	Subject Likes	Class Average	Student Marks	Prediction
Sebolt	76.36	ElecTrade	78.18	63	60	0
Lalitha	67.27	Maths	81.82	53.88	62	0
Thomas	71.43	IndElec	73.64	68.04	83	1
Lalitha	67.27	EngSc	75.45	61.13	69	0

This algorithm works on the premise that students can first be classified in date bands, as far as their dates of birth are concerned, and then a prediction of the probability of that student to succeed or fail in the academic environment can be made. The following is a WEKA API that is appropriate to use in conjunction with java in the development of the model.

The accuracy of the Decision Tree in this case yields a result of 90% when there was a 70% split of the training data.

VI. Conclusion

The target variable, c_i, in this case what the student average pass mark will be above 80% or not. The attributes used in this decision tree are as follows:

- x_1 =Lecturer
- x_2 = Lecturer Likes
- y_1 = Subject
- y_2 = Subject Likes

And the prediction class variable is as follows:

- c_1 = Student Marks ({0, 1})

This paper has presented the decision tree algorithm as one of the algorithms to predict success or failure for students in FSASEC - VUT. The implementation of this algorithm yielded 90% accuracy. It is therefore concluded that the decision tree algorithm can be incorporated in the Academic Environment Model to assist lecturers and management to make informed decisions about student performance in

FSASEC. Java and the WEKA API can be used to implement the prediction tool in order to improve selection of students, identification of those at risk and placement of top performing ones for more opportunities.

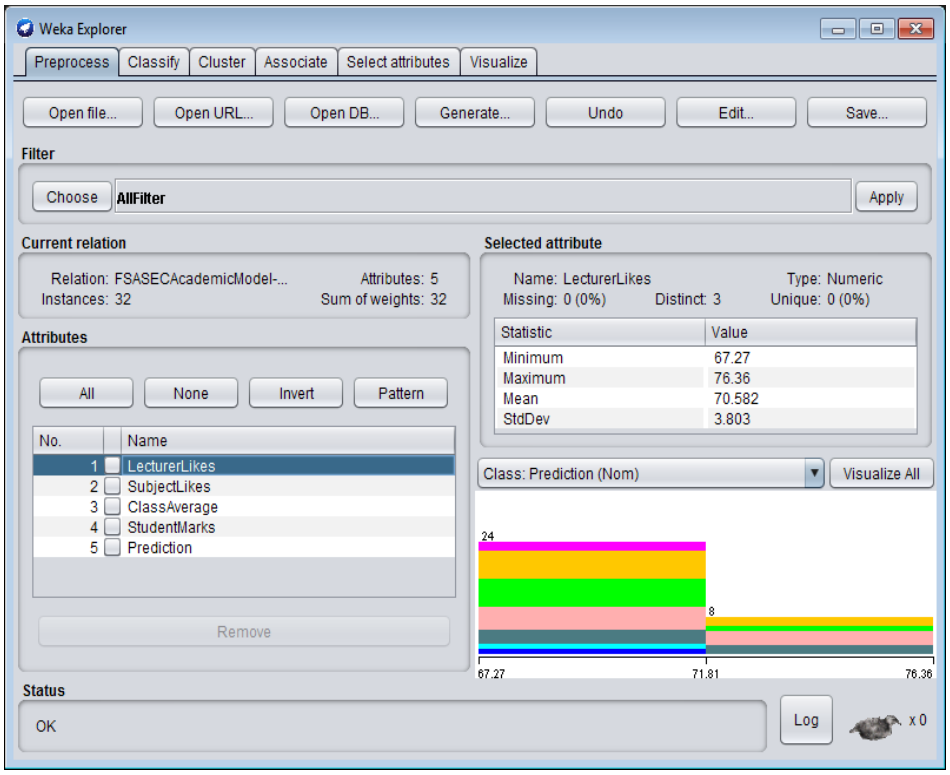

Fig. 6 Academic Environment Model Using WEKA

REFERENCES

[1] T. Daniel Larose, D. Chantal Larose, "*Data Mining and Predictive Analytics"* Second Edition, Wiley. 2015.
[2] R. Paul and L Cowe Falls. "*Mapping Career Success Competencies to Engineering Leadership Capabilities*, 2015 IEEE.
[3] M. Agaoglu. "*Predicting Instructor Performance Using Data Mining Techniques in Higher Education*" 2016 IEEE.
[4] J. Han,M. Kamber and J. Pei, "*Data Mining"* Third Edition, Morgan Kaufmann Publications. 2012.
[5] Ministry of Education. "*A New Funding Framework: How Government grants are allocated to Higher Education Public Institutions*" February 2004.
[6] A. Giri, M. V. V. Bhagavath, B. Pruthvi and N. Dubey. "*A Placement Prediction system Using K-Nearest Neighbor Classifier"* 2016 IEEE.
[7] B. Twala. "*Robot Execution Failure Prediction Using Incomplete Data"* Proceedings of the 2009 IEEE International Conference on Robotics and Biometrics, December 19-23, 2009, China.
[8] P. K. Imbrie and J. Lin. "*Work in Progress Engineering Students Change in Profile over the Freshman Year across Male and Female Samples: A Neural Network Approach"* 36th ASEE / IEEE Frontiers in Education Conference.
[9] S. Fong, Y. Si and R. P. Biuk-Aghai. "*Applying a Hybrid Model of Neural Network and Decision Tree Classifier for Predicting University Admission"* 2009 IEEE.
[10] B. Twala and T. Nkonyana. "*Extracting Supervised Learning Classifiers from Possibly Incomplete Remorely Sensed Data*" 2013 BRICS Congress on Computational Intelligence. 478 – 479 Brazil, 2013.
[11] W. Shang and H. Zhu. "*The Improved ontology kNN Algorithm and its Application*" 2006 IEEE.