Journal homepage www.jzs.uos.edu.krd
Journal of Zankoi Sulaimani
Part-A- (Pure and Applied Sciences)

Classification Performance Enhancement for Students Realisation Model

Tarik A. Rashid

Software Engineering, College of Engineering, Hawler, Kurdistan, Iraq

E_mail : tarik.rashid@su.edu.krd

Article info

Original: 20 Apr. 2015
Revised: 6 June 2015
Accepted: 25 June 2015
Published online:
20 Sep. 2015

Key Words:

Forecasting Student Performance, Feature Reduction, Neural Networks, Support Vector Machines, K-Nearest Neighbors, Genetic Algorithms.

Abstract

This research work aims at enhancing a classification task for student's realisation model at Salahadin University, Hawler, Kurdistan. 1000 records of data from different colleges and departments at Salahadin University are collected to conduct this research work. The collected data has been pre-processed, cleaned, filtered, normalized, then after, feature selection techniques are applied to reduce the dimensionally of the data, finally a classification task is carried out to find the realization of students. The results show that a model of Support Vector Machine +Genetic Algorithm + Artificial Neural Network produces promising results than other models.

Introduction

Usually, academic universities register students in admission under varied courses from completely different places, backgrounds, and score values. Besides, schools, faculties and departments could also be related to completely different boards, an ordinary board can have entirely different subjects in their curricula and furthermore can have different levels of magnitude in their subjects. Investigating the past performance of admitted students would offer a stronger perspective of the probable educational performance of scholars within the future. This can very well be achieved through the concepts of data mining and machine learning ***[1]***.

Obviously, any education quality in university is solely reflected in its analysis work and educational activities, thus, it's therefore right to mention that the amount of accepted students can have an effect on the level of the colleges ***[2]***.

It is very helpful for any education system to own its correct student performance system. The correct student performance system permits body workers to be ready to recognise between accepted and not accepted candidate students for any educational course or subject. The absence of getting a flourishing system to perform exactly prediction results might influence unsuitable students to the course or subject.

It's therefore important to own student performance system to make correct prediction or to create right selections so that to assist educational stuff improve student educational performance. Also facilitating further aid and support for college students such as custom help and tutoring resources or perhaps the lecturers are going to be able to establish the most satisfactory education engagements for each single bunch of student, plus, the lecturers are going to be able to supply their help through made-to-order substantial materials to students.

Consequently, good valuation of student action possibly will be a smart mechanism to revitalise the students' educational level and supply higher academic services. As a result, the power to predict students' educational performance is very important for academic establishments. It is worth pointing out that the goal may well be achieved via the use of machine learning and data mining techniques. These techniques have great ability to process huge data, extract hidden patterns and discover significant relationships so that to be able to make decision ***[1]*** and ***[2]***.

Data mining supported by machine learning and soft computing techniques are establishing a new field of research work named educational data mining ***[3]*** and ***[4]*** at universities levels. Authors in ***[5]*** stated that new and useful knowledge about students can be discovered via the application of data mining in education.

Authors in ***[6]*** stated that techniques for exploring the types of data that come from educational institutions can be developed via educational data mining. There are various techniques of data mining, examples of these statistics, visualization, clustering, classification, and revealing of outliers. Among these, classification is one of the most frequently studied techniques. Classification may possibly be a method of supervised learning wherever information is separated into totally different categories. Classification maps information into pre-arranged groups of types. The key objective of a classification model is to envisage the target class for every sample within the dataset. There are numerous techniques for classification of data, these are namely support vector machine (SVM), artificial neural network (ANN), and Bayesian classifier ***[6]***.

According to above techniques, the classification task can be conducted via describing and distinguishing data categories. After that the designed classification model is employed to foresee the group label of recent data which does not fit with the trained dataset. Clearly these classification techniques have been used commonly in educational settings ***[7]*** and ***[8].***

There are some software computing techniques such as ANNs and fuzzy set theory; these are good techniques for establishing intelligent models. An IF –Then rule concept can be employed through a fuzzy inference system (FIS) technique, it is a way of obtaining knowledge from human experts that can work out inaccurate and ambiguous difficulties ***[9]***. The idea of fuzzy inference system has been commonly employed in various tasks such as control, system identification, and furthermore, these ideas are used to solve optimisation problems. The different between ANNs and FISs is that ANNs normally learn via updating its connection weights, usually ANNs have the ability to learn from either from their environment or they get self-organized, and can adapt in a cooperative way, whereas, FISs normally lack learning abilities.

Thus, for the above motive, an amalgamation of ANNs and FISs has been presented and called a neuro-fuzzy system. Authors in ***[10]*** have produced a complete fuzzy-rule-based system. Obviously, the values of ANNs and FISs can be integrated in a neuro-fuzzy approach. Essentially, neuro-fuzzy systems defined as fuzzy networks that comprise ANNS and FISs, nonetheless, this idea can also tackle some downsides of ANNs, plus, the drawbacks of FISs since they are able to learn and have ability to represent knowledge in a reflection way and great learning aptitude ***[11]*** and ***[12]***.

A neuro-fuzzy classifier is another technique of neuro-fuzzy systems, which amalgamates the dominant sketch of FISs with ANNs learning abilities so that to divide a feature space into groups. A neuro-fuzzy classifier has been extensively employed to solve different applications . Authors in ***[2]*** managed to employ a neuro-fuzzy classifier with a technique called a scaled conjugate gradient so that to enhance the work that

were carried out via Cetişli and Barkana ***[12]*** to perform a classification task for students depending on levels of their expected academic performance.

This research work aims at using ANNS as a classifier, plus K-nearest neighbor with genetic algorithm and SVM with genetic algorithm are suggested for feature reduction to predict student's performance at the end of academic year.

In this paper, artificial intelligence techniques with some feature selection techniques are used as classifiers to predict student's performance at the end of their course. The aim of this study is to enhance the classification task so that to obtain accurate prediction for students' performance. The structure of this paper is as follows: in the next section, the structure of design method is explained, after that, details of data collection, cleaning and preparation are given, then, dimension reduction techniques are explained, next classification model is detailed, later, experimental results are described, finally, the main points in this research work are concluded.

Design and Methods

The system structure of student performance model is shown in *Figure 1*: Data collection, Dimension Reduction, Classification Model and results. The following sections are the details description of all phases of the system.

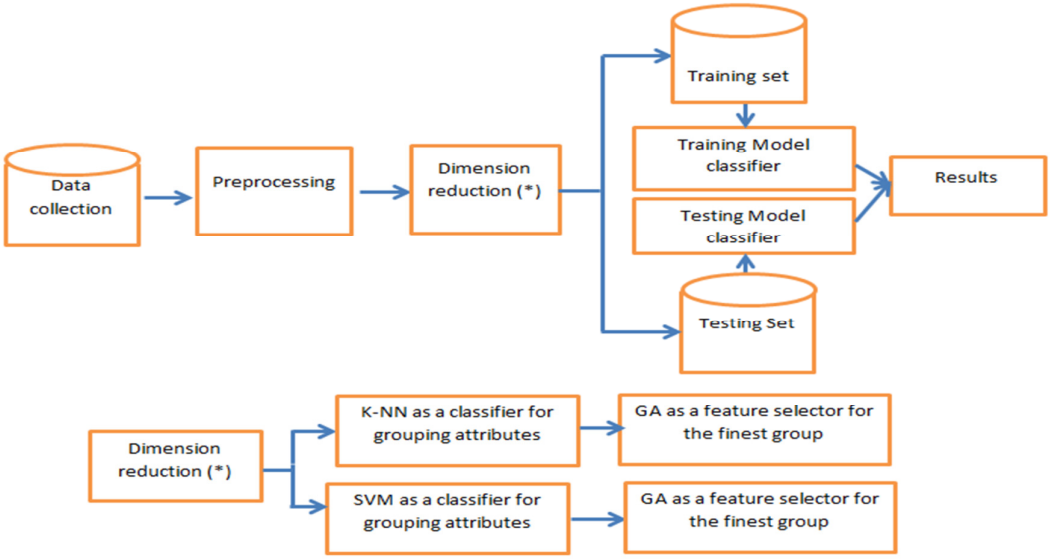

Figure -1: Student Performance Model.

A. Data collection

Data assortment contains 2 styles of data, these are namely; primary and secondary knowledge. First, primary data, Questionaries' is conducted to gather the first primary data from the students. The effects within the form comprised of the students' demographic data on age, gender, parents education level, high school type, address, scores at national examination, score of English at the final year of high school, score for university English module, and general test for this study are considered.

Additionally, the students were furthermore expected to incorporate their colleges and departments, since there was a demand for this information to allow connection to the secondary data of the study. For the

secondary data, this knowledge is collected from university papers and forms including admission and educational performance records of every student who filled the questionnaire form.

B. Data Size

There are around 4000 students at the first year, at the university. In general, students are registered for this program is from different departments such as science, engineering, education, humanities and others. Not to mention that the students are coming from different backgrounds with various previous achievements. In this research work, a population of 1000 student's records has been nominated.

C. Data Factors

In this research work, the data is collected and prearranged to comprise examples with detail of different factors which affect student's performance. The reason behind this selection was because the study mostly focuses on the socio-economic background and tutors expertise. The enrollments at these departments require close National Exam Score but students have variety of social backgrounds. The data collection can be more elaborated in the following subsections:-

D. Input variables and Output variables

Over extensive search of the past works, a number of socio-economic, academic, and other related factors that are considered to have influence on the performance of a university student have been identified. Many researchers conducted detailed studies about the factors contributing student performance at different study levels and for different study programs. In this study, the following factors have been taken into consideration, Gender numeric, age numeric, personal address city, personal address town, Education of Degree of Father, Education of Degree of mother, High school address city, High school address town, High school address village, High school type, High school language, Overall score at national exam, English score at national exam, College, Department, English Tutor, English Tutor English Local, English Tutor Native, English Module Score for year one (Department Exam), English Module Score for year one (General University Test)-start of course. The outputs are values for English Module Score for year one (General University Test)-End of course class, the output is either pass or failed.

E. Data Pre-Processing and Preparation

This is a crucial step for any classification or prediction model. It permits finding any uncommon knowledge, distinguishing different variables, detecting invalid cases and incorrect variables in the data set itself. There is no doubt that the accuracy of the model extremely relies on the correctness of the information and clean data result. During this study, knowledge validations for all knowledge classes, knowledge nominal values and knowledge integrity are checked and clean consequently. The factors described previously have been converted to input variables by normalisation them depending on their available domains.

As mentioned earlier that initially the students data is collected via a constructed questionnaire. Next the normalisation process is conducted in the pre-processing phase so that to prepare for processing. The process of data cleaning is carried out via the parameters used in the data analysis and obviously the data that are missing are removed or completed via null values. The process of diagnosing variables is conducted in this phase or can be carried out in different process. Data integrity is accomplished via examining the data against the values and data types and their ranges. The last step in this phase is to segregate data into three different categories namely; training, validation and testing data after cleaning phase, 650 records met the requirements and used in the study. Finally, the collected data has been fragmented randomly into training set (80%) and testing set (20%).

F. Dimension Reduction

In machine learning and statistics, dimensionality reduction or dimension reduction is the process of reducing the number of random variables in a data set. This part is consisted of two sections; these are namely grouping and assortment. These can be explained as follows:

1. Grouping: in this stage the data is taken and divided into different groups based on the similarity of features in the data set, K-Nearest Neighbor and Support Vector Machine techniques are used independently to segregate data into different groups. Following subsections are the descriptions of both techniques (items I and II).
2. Assortment: This stage comes after the grouping stage, the main purpose of this stage is to choose the ultimatum group among all the groups of data in stage one (as in item II).

I. K-Nearest Neighbor

It is preferable to take into account each of the features in the training set as a distinct dimension in some space, take the value an observation has for this characteristic to be its coordinate in that dimension, therefore obtaining a group of points in space. Then, the similarity of two points is taken into account as the distance in this space under some acceptable metric ***[13], [24]*** and ***[15]***.

The approach within which this technique decides which of the points from the training set are similar enough to be considered when selecting the category to predict for a new observation is to choose the K nearest data points to the new observation, and to take the foremost common category among these. The technique K-Nearest Neighbor (K-NN) may be described as: first, K is indicated as a positive number, in conjunction with a new example. Second, the K entries that are closest to the new example are chosen from the data set. Third, the foremost common classification of those entries will be found. And finally, this is the classification we have a tendency to provide to the new sample.

Given *N* training vectors, K-NN algorithm identifies the *K* nearest neighbors of *'C'* in spite of labels; two classes are A and B. If we assume K=3, then we find classes for *C*. if *K=1* each training vector is considered as a region in space. This is defined by Voronior partition space as shown in equation (1).

$$R_i = \{x, d(x,x_i) < d(x,x_j) \qquad where \qquad i \neq j \tag{1}$$

II. Support Vector Machine

The idea of Support Vector Machine is widely implemented for classification tasks. This idea can be applied simply via taking inputs from a feature vector and estimating a class for all sets of inputs. Therefore, the dimension of the feature vector is mostly depending on the number of extracted features. In the literature, support vector machines are mainly used by the researchers for classification tasks and they are measured to be a problem of a quadratic optimization. It is concluded that the most important function of support vector machines is to dwindle the error between inputs and desired outputs ***[16]*** and ***[17]***. Various data and information classes in the feature space can actually be separated by support vector machine through having the maximum hyper plane which can either be described as a line in two dimensions or a plane in more than two dimensions. These hyper planes are normally functioning in order to amplify the support vector machines' oversimplification ability. Thus, the classifier will produce an optimum hyper plane, this means, searching the maximum margin hyper plane between two or more classes is the crucial goal ***[18]*** and ***[19]***.

III. Genetic Algorithms

The technique of Genetic Algorithms (GA) is considered as accommodative heuristic search algorithm supported the organic process concepts of survival of the fittest in Biology. It denotes an intelligent use of a random search which is accustomed to solve enhancement issues. However, GA uses historical data to guide the search into a higher performance region among the search area. This technique is essentially intended to

mimic processes in natural systems necessary for evolution; mostly those follow the principle of survival of the fittest that coined by natural scientist Charles Darwin.

Naturally speaking, struggling among individuals for useful resources tips to controlling the stronger people over fragile ones. In other words, as mentioned, this technique solve problem by imitating the concept of the survival of the fittest among people over consecutive generation ***[20]***.

In every single generation there is a population that consists of strings of characters which are corresponding to the chromosome that can be found in human DNA. It has been said that every single individual reflects a point in the searching space and it is a potential solution for the problem. Then, a process of evolution will make these individuals in the population go over and done with the process. Genetic algorithm is basically an example of genetic structure and chromosomes behavior inside a population of individuals. Some important features about GA can be described as follows: first, in a population, the individuals are contending for resources and companions; second, individuals that are prosperous in a rivalry can generate additional descendants than individuals that perform badly; third, worthy individuals proliferate inheritable factor all the way through the population in order for two worthy fathers will sometimes produce descendants that are higher than either parent. And finally, accordingly every single progressive generation can become better fitted to their atmosphere ***[20]***.

G. Classification Model

Artificial neural networks these are either supervised or unsupervised. The network in the supervised learning mode is trained via a set of real data which is called the training set. Plus, the real output or target for every single input sample is made available to the network during the training. The inputs are processed and then the processed output results are compared against the desired targets so that to get the network errors which in return are back propagated through the layers of the network. The back propagated error can cause the network to adapt its weight connections via which the network can be controlled ***[21]***.

On contrary, in unsupervised learning, the targets are not included in the training and testing sets, in other words, there are not targets at all, thus only the input features in the training samples are delivered to the network.

The supervised learning process in this research study is employed, since the performance of the students at the end of the course is known after sitting for their final exams in the course. The final exam of the course is used as a target for the neural network.

The data set in this work is divided into three sub sets; these are namely, training, testing, and validation sub sets. First the training set is employed to detect the inputs and outputs mapping via updating the connection weights. In this study, 650 samples are used, 80% of this study is used as a training set and the 20 % of the data is used as a testing set. The degree of network learning is checked via using the verification set so that to find out whether the network is come together correctly for sufficient generalisation capacity or not ***[22, 23]***.

Experimental Results

Several experiments in terms of different network structures, parameters, and features selections were carried out. Three main experimental models are selected; these are namely; ANN, KNN+GA+ANN and SVM+GA+ANN. The first model is ANN which uses no feature selection technique, the second model uses K-NN for group the feature inputs based on their similarities, then it uses genetic algorithm to select the finest group and the last model which is SVM+GA+ANN, uses SVM for grouping the features and GA for selecting the finest group. *Table1*: shows the three different models with details of mains parameters for each structure.

As can be seen in ANN model all features are used and the number of hidden neurons is 7, and the training cycles is 800, however in model KNN+GA+ANN, 3 hidden neurons and 400 training cycles are used and the number features are reduced to 7 (F4,F5,F12,F13,F16,F17,F19), this is because of the process

of feature reduction which is caused by using K-NN and GA, thus the features are reduced from 20 to 7 features.

Table - 1: shows the NN Model, KNN+GA+ANN Model and SVM+GA+ANN

Models	***ANN***	***KNN+GA+ANN***	***SVM+GA+ANN***
Learning rate	*0.01*	*0.3*	*0.2*
Momentum	*0.7*	*0.7*	*0.3*
Training Cycles	*800*	*400*	*1000*
Number of hidden neurons	*7*	*3*	*10*
Network stop condition	*0.001*	*0.001*	*0.001*
Training set	*80%*	*80%*	*80%*
Testing set	*20%*	*20%*	*20%*
K nearest neighbors	-	*1 or odd for two classes*	-
Kernel function		-	*Polynomial*
GA Population size	-	*20*	*20*
GA Number of generations		*20*	*20*
GA Probability of crossover		*0.6*	*0.6*
GA Probability of mutation		*0.033*	*0.033*
Number of Features	*20 (all features)*	*7 features: (F4,F5,F12,F13,F16,F17,F19)*	*15 features: (F2,F3,F7,F8,F9,F10,F11, F12,F13,F14,F15,F16,F17 ,F19,F20)*

In SVM+GA+ANN model the number of hidden neurons is 10, the number of training cycles is 1000 and, the number of feature is reduced to 15. (F2,F3,F7,F8,F9,F10,F11,F12,F13,F14,F15,F16,F17,F19,F20), this is because of the process of feature reduction which is caused by using SVM and GA, thus the features are reduced from 20 to 15 features.

Figure 2: shows the performance of SVM+GA+ANN, KNN+GA+ANN and ANN, obviously the classification is improved via using SVM+GA+ANN.

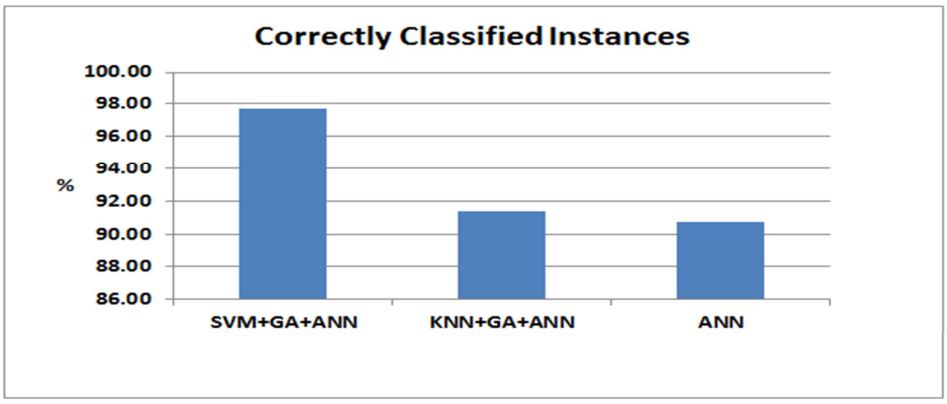

Figure -2: shows correctly classified instances of SVM+GA+ANN, KNN+GA+ANN and ANN models.

Figure 3: shows the performance of SVM+GA+ANN, KNN+GA+ANN and ANN, the incorrectly classified instances is drastically reduced in SVM+GA+ANN, and KNN+GA+ANN.

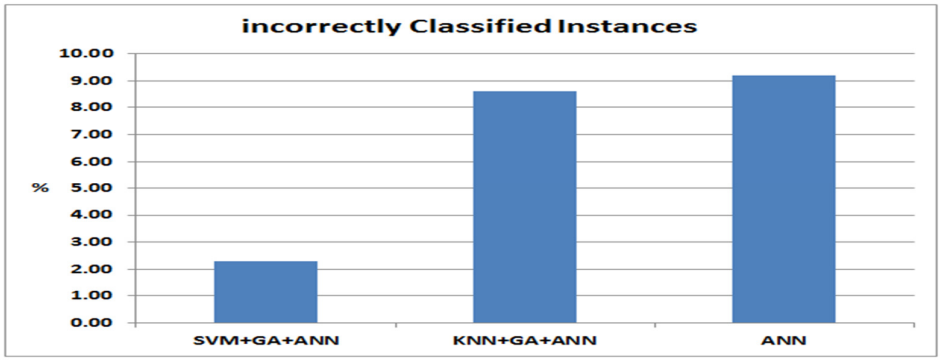

Figure -3: shows incorrectly classified instances of the three models.

Figure 4: shows that model SVM+GA+ANN produces smallest mean square error.

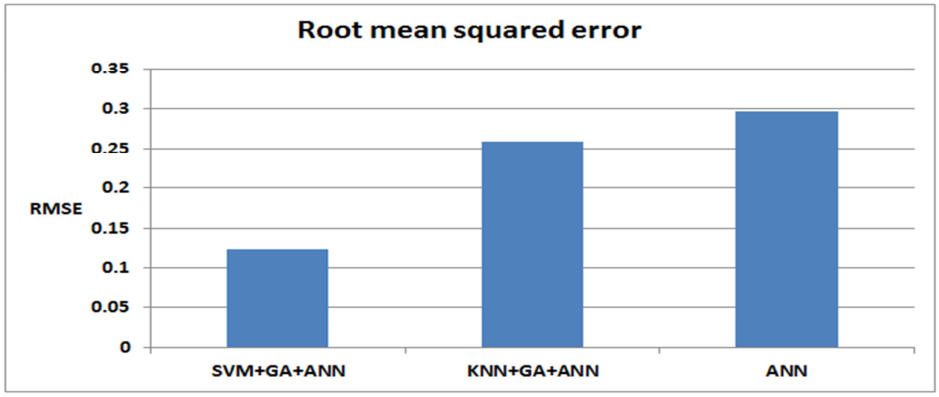

Figure -4: shows root mean squared error.

Figure 5: shows overall performance for the three models, SVN+GA+ANN produces better performance with 15 features.

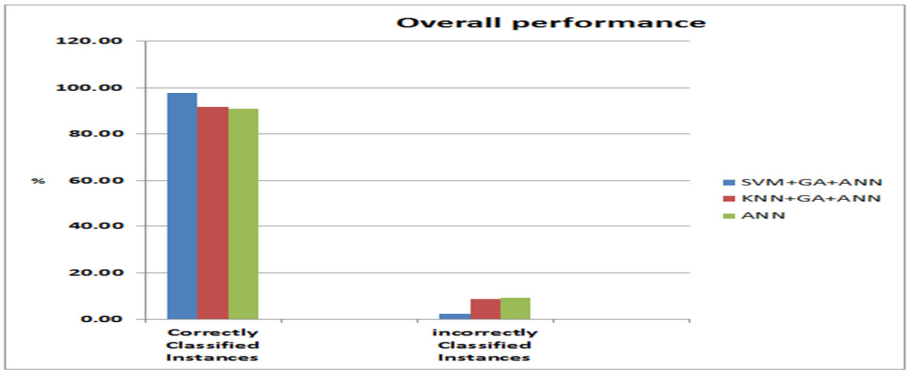

Figure -5: shows overall performance of all models.

Figure 6; shows the performance of the three different models in terms of time. It is clear that SVM+GA+ANN model spent the longest time for classification, this is due to the number of features and plus, the number of hidden neurons which is 10, plus the number of training cycles, this means the ANN takes more time to process hidden neurons, in addition, the number of features is relatively large (15). However the saving grace about SVM+GA+ANN produces the most accurate results, of course this is at the expense of time.

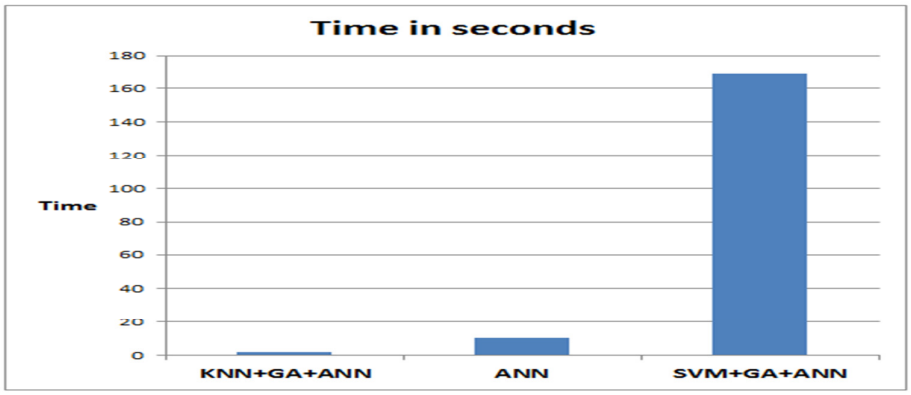

Figure -6: shows performance of the three models in terms of time.

Conclusion

Increasing technologies with the speedy economic process have impacted our lives generally and education system particularly, thus, there is a great demand by students for studies in Kurdish Universities in Kurdistan has been emerged. Consequently, this can impact the action of target for any new established programs at universities. It's so essential to possess student performance system and create correct prediction to help academic stuff improve student academic performance. This paper meant to enhance a classification task for student's realisation model at Salahadin University. A new data set is collected from faculties of science and education; the collected data is pre-processed, cleaned, and analysed. A new technique is employed for feature selection. SVM and KNN on an individual basis are used to segregate data features into groups that have similarities in terms of distances. Then after, GA is employed to choose on the best group of features. Finally, a classification task is performed using Artificial Neural Network. The results show that the combined techniques like SVM+GA+ANN and KNN+GA+ANN are produced better results than simply ANN. Some important aspects about KNN is also observed such as choosing *K* as an odd number for two classes, and *K* must not be multiple of the number of classes, we have realised the downside of KNN is the complexity of searching the nearest neighbors for each example. Thus, features that were selected by SVM are more reliable to get more accurate results.

References

[1] Adhatrao K. Gaykar A. Dhawan A., Jha R., Honrao V. *" Predicting Students' Performance Using Id3 And C4.5 Classification Algorithms*", International Journal of Data Mining & Knowledge Management Process (IJDKP). Vol.(3), No.5, (2013).

[2] Hung Do Q. and Chen J. F. *"A Neuro-Fuzzy Approach in the Classification of Students' Academic Performance"*. Computational Intelligence and Neuroscience, Academic Editor: Daoqiang Zhang. (2013).

[3] Merceron A. and Ycef K. "Educational data mining: a case study" in *Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED '05),* IOS Press, Amsterdam, The Netherlands (2005).

[4] Romero C. and Ventura S. *"Educational data mining: a survey from 1995 to 2005"*, Expert Systems with Applications, vol. (33), No.1, 135–146 (2007).

[5] Barker K., Trafalis T. and Rhoads T. R. *"Learning from student data*", in Proceedings of IEEE Systems and Information Engineering Design Symposium, pp. 79–86 (2004).

[6] Sharma A., Kumat R., Varadwaj P. K., Ahmad A., and Ashraf G. M. *"A comparative study of support vector machine, artificial neural network and bayesian classifier for mutagenicity prediction"*. Interdisciplinary Sciences, Computational Life Sciences, Vol. (3), No. 3, pp. 232–239 (2011).

[7] Bhardwaj B. K. and Pal S. *"Data mining: a prediction for performance improvement using classification"*. International Journal of Computer Science and Information Security, Vol. 9, No. 4, pp. 1–5 (2011).

[8] Huang S. and Fang N. *"Predicting student academic performance in an engineering dynamics course: a comparison of four types of predictive mathematical models"*. Computers & Education, Vol. 61, pp. 133–145 (2013).

[9] Norazah Y., Nor B. A., Mohd S. O. and Yeap C. N. *"A concise fuzzy rule base to reason student performance based on rough-fuzzy approach"*. In Fuzzy Inference System-Theory and Application, InTech (2010).

[10] Ata R, and Kocyigit Y. *"An adaptive neuro-fuzzy inference system approach for prediction of tip speed ratio in wind turbines"*. Expert Systems with Applications, Vol. (37), No. 7, pp. 5454–5460 (2010).

[11] Singh R., Kainthola A. and Singh T. N. *"Estimation of elastic constant of rocks using an ANFIS approach"*. Applied Soft Computing Journal, Vol. (12) , No. 1, pp. 40–45 (2012).

[12] Cetişli B. and Barkana A., *"Speeding up the scaled conjugate gradient algorithm and its application in neuro-fuzzy classifier training"*. Soft Computing, Vol. (14), No. 4, pp. 365–378(2010).

[13] Mirkes E. "KNN and Potential Energy (Applet)". University of Leicester. Available: http://www.math.le.ac.uk/people/ag153/homepage/KNN/KNN3.html /16/02/2015 (2011).

[14] Sergios T. and Konstantinos K. *"Pattern Recognition",* Second Edition Published by Academic Press (2003-04-09) ISBN 10: 0126858756 / ISBN 13: 9780126858754 (2003).

[15] Kozma L. K. *" Nearest Neighbours Algorithm"*. Helsinki University of Technology, Available: http://www.lkozma.net/knn2.pdf, /01/02/2015 (2008).

[16] Chih-Wei Hsu, Chih-Chung Chang and Chih-Jen Lin. *"A Practical Guide to Support Vector Classification"*. Department of Computer Science, National Taiwan University, Taipei, Taiwan (2010).

[17] Chih-Chung Chang and Chih-Jen Lin. *"LIBSVM - A Library for Support Vector Machines"*. Department of Computer Science, National Taiwan University, Taipei, Taiwan,.URL: http://www.csie.ntu.edu.tw/~cjlin/libsvm/(2013).

[18] Rashid T. A. and Abdul Hamid S. *"Support Vector Machines for Predicting Electrical Faults"*, in the International Journal, University Technology, Baghdad, Iraq, Eng. & Tech. Journal ,Vol.32, , No.8, Part (A), (2014).

[19] Ali A. M., Md Jan Nordin and Azizi Abdullah *"A Spatial Visual Words of Discrete Image Scene for Indoor Localization"*. Research Journal of Applied Sciences, Engineering and Technology, ISSN: 2040-7459; e-ISSN: 2040-7467 © Maxwell Scientific Organization Vol. (7) No, 14, pp. 2806-2812 (2014).

[20] Melanie M. *"An Introduction to Genetic Algorithms*". A Bradford Book The MIT Press Cambridge, England (1999).

[21] Huang B.Q. , Kechadi T.-M. , Buckley B., Kiernan G. , Keogh E. , Rashid T. , *"A new feature set with new window techniques for customer churn prediction",* Expert Systems with Applications, Expert Systems with Applications, Vol. (37) pp. 3657–3665, (2010).