Prediction of Student Academic Performance using Neural Network, Linear Regression and Support Vector Regression: A Case Study

Efrem Yohannes Obsie
Beijing Jiaotong University
School of Computer and Information Technology
and Hawassa University Institute of Technology

Seid Ahmed Adem
Beijing Jiaotong University
School of Computer and Information Technology
and Debre Berhan University School of Computing

ABSTRACT
Predicting students' academic performance is very crucial especially for higher educational institutions. This paper designed an application to assist higher education institutions to predict their students' academic performance at an early stage before graduation and decrease students' dropout. The performance of the students was measured based on cumulative grade point average (CGPA) at semester eight. The students' course scores for core and non-core courses from the first semester to the sixth semester are used as predictor variables for predicting the final CGPA8 upon graduation using Neural Network (NN), Support Vector Regression(SVR), and Linear Regression (LR). The study has verified that data mining techniques can be used in predicting students' academic performance in higher educational institutions. All the experiments gave valid results and can be used to predict graduation CGPA. However, comparisons of the experiments were done to determine which approaches perform better than others. Generally, SVR and LR methods performed better than NN. Therefore, we recommend the adoption of SVR and LR methods to predict final CGPA8, and the models can also be used to implement Student Performance Prediction System(SPPS) in a university. Thus, the study has used the models from SVR and LR methods for designing an application to do the prediction task.

Keywords
Educational Data Mining, CGPA, Linear Regression, Neural Network, Support Vector Regression, Student Performance Prediction System

1. INTRODUCTION
Data mining can discover hidden information to inform decision-making in various domains. Data mining has a wide range of applications in different areas, including marketing, telecommunications, scientific discovery, surveillance, banking, fraud detection, and educational research[1].The education system is one of these domains in which the primary concern is the evaluation and, in turn, enhancement of educational organizations.

The availability of educational data has been growing rapidly, and there is a need to analyze huge amounts of data generated from this educational ecosystem, as a result, Educational Data Mining (EDM) field has emerged. Educational data mining is the method of applying data mining tools and techniques to analyze the data at educational institutions[2]. EDM is relatively new field of data mining applications emerged in 2005, EDM is concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. EDM develops methods and applies techniques from statistics, machine learning, and data mining to analyze data collected during teaching and learning[2].

Universities have been using many data mining techniques to analyze educational report stored in the educational institute such as enrollment data, students' performance, teachers' evaluations, gender differences, and many others. Data mining techniques may, for example, give a university the needed information to better plan a number of students' enrollment, students' dropout, early identification of weak students, and to efficiently allocate resources with a precise approximation. Data mining is a powerful tool for academic intervention. Through data mining, a university could, for example, predict with 85 percent accuracy which students will or will not graduate. The university could use this information to concentrate academic assistance on those students most at risk.

Students' low academic performance at the end of a university degree has been a long-standing problem, especially in developing country. Today, universities are working in a very dynamic and powerfully viable environment. Hence, they gather large volumes of data with reference to their students in electronic format. However, they are data rich but information poor which results in unreliable decision making. The main challenge is the effective transformation of large volumes of data into knowledge to improve the quality of managerial decisions and to predict academic performance of students at an early stage in order to help universities, teachers not only focus on bright students but also to initially identify those students with low academic achievement and find ways to support them [3].

This paper presents a prediction of students' academic performance from educational database particularly using their scores, with no economic, social and psychological factors. From a managerial point of view, to gather marks of students from the educational database is much easier than gathering students economic, social and psychological factors using questionnaire and interview. Thus, if a reasonable prediction can be reached with scores only, it makes the implementation of a Student Performance Prediction System (SPPS) in a university easier[4]. Furthermore, we developed Neural Network(NN), Support Vector Regression(SVR) and Linear Regression(LR) models to predict students' academic performance which is measured by the students' final CGPA8 (CGPA at semester eight).

2. LITERATURE REVIEW

In this section, literature related to student academic performance prediction are reviewed.

In [5] the authors adapted the methodology to be used for a small dataset, 48 students enrolled in engineering dynamics course. They developed and tested two algorithms namely: Support Vector Machine(SVM) and Multiple Linear Regression(MLR). The three criteria for model evaluation includes accurately predicted grade range, maialami, and missing alarm. The results examined showed that SVM model produces higher accuracy to identify students having low grades.

In addition, the authors in [6] analyzed students' performance data using classification algorithm named ID3 to predict students marks at the end of the semester. This was applied for the master of computer applications course from 2007 to 2010 in VBS Purvanchal University, Jaunpur. Their study aimed to help students and teachers find ways to improve students' performance. Data were collected from 50 students, and then a set of rules was extracted for their analysis.

Another study in [7] designed a tool using the .NET framework to predict students grade by providing input parameters. Models based on the students' enrollment data were developed using ten classification trees (OneR, Random Forest, ZeroR, Random Tree, Decision Stump, REPTree, JRip, J48, PART, and Decision table) and a multilayer perceptron learning algorithm by using WEKA. A framework is built for intelligent recommender system which recommends suitable action for improvement. The work is based on the background factors that predict the tertiary first-year academic performance of students. The data for the study is taken from Babcock University, Nigeria. The background factors for the students were collected through in-depth interview. The various demographic factors are father occupation, mother occupation, family income, place of birth, family size, academic qualification of parents, parents' marital status. The benchmark used in the comparison of the generated models include confusion matrix, accuracy and time. Random tree outperformed the other algorithms in terms of the benchmark. Therefore, Random Tree is adopted as the best performing algorithm in the domain of the study to serve as building block for designing the generic system.

The researchers in [8] use those factors to build prediction model which has not been used by anyone so far. Various classification techniques are used on students' social integration, academic skills, and emotional skills. The data is collected from Guru Gobind Singh Indraprastha University for Master of Computer Application students, to predict their performance in the third semester. Two algorithms were used: J48 and Random Tree for early prediction. The reason to consider the third semester is most of the students are observed to drop out after their first year. In addition, the students normally take a year to integrate into the environment of any academic institute. The authors aim is to study the impact of various factors in predicting the performance of the students. It has been found that the result of the second semester strongly influences the result of the third semester, especially the programming courses.

The performance of various classifiers was compared using educational data mining as in the following research. The study in [9] aimed to predict students' enrollment using admissions data. The researchers used applicants' data from West Virginia University that consists of 112,390 instances. Various classification learners' models had been built. They compared the result of the different learners and identified that the rules from J48 and Rido to be the best.

Moreover, in [10] the authors attempted to apply different classification techniques to an educational dataset to compare their performance and choose the best algorithms to be integrated into their (E-learning Web Miner) tool. This tool aimed to help teachers discover their students' performance. They used the data from the course named "Introduction to multimedia methods", offered in three academic years from 2007 to 2010 at the University of Cantabria. They had used different classification methods and they found that the performance and the accuracy of the techniques rely on the type of the attributes and the size of the dataset. Among the findings, J48 was found suitable for datasets with more than 100 instances and nominal attributes with missing data.

In a study of advanced programming course in an institute of higher learning in Malaysia in [11], the authors have presented a theoretical model that shows how data from different educational settings can contribute in the prediction of student's final grade. The results indicate that coursework marks have the most significant positive relationship with the student's final grade followed by a total number of materials downloaded from course management system.

Furthermore, the research paper in [12] finds that performance in the first year of Computer Science courses is a determining factor in predicting students 'academic performance at the conclusion of the degree. They consider the data of 85 students in the School of Computing and Information Technology at the UTECH and analyze this single cohort of students through the entire degree. They find that the first-year gateway courses like C Programming, Introduction to Computer Networks and Computer Logic & Digital Design are strong predictors for overall academic performance (Grade Point Average GPA) in BSCIT program at UTECH. They use statistical methods like regression, no other data mining classifier, and find a strong correlation between the performance in first-year Computer Science courses and students overall performance in BSCIT program with a correlation of 0.499 that explains 70.6% of students 'performance. The authors also concluded that students 'demographics do not have any significant relation to academic performance.

The paper in [13] used the Adaptive Neuro-Fuzzy Inference System (ANFIS) for student academic performance prediction. The proposed approach consists of two steps. First, results of the students in the previous exams are preprocessed by normalizing the results in order to improve the accuracy and efficiency of the prediction. Second, the ANFIS is applied to predict the students' expected performance in the next semester. Three ANFIS models: ANFIS-GaussMF, ANFIS-TriMF, and ANFIS-GbellMF that utilized various membership functions to generate accurate fuzzy rules for predicting the student's performance were used. The experimental results showed that ANFIS-GbellMF model outperformed the other ANFIS models with a Root Mean Square Error (RMSE) as low as 0.193.

3. PROPOSED WORK AND METHODOLOGY

3.1 Proposed Approach

The first and the foremost step is to collect the dataset required for the study. The methodology is applied to a factual data having information about the students who did their graduation from School of Computer Science, Hawassa

University, Ethiopia. Once the data is gathered, next step is to transform into the required format to make it ready for the mining process, which is known as a pre-processing phase. It is a crucial step in data mining systems that aspire to transform the raw data into a proper format for resolving a particular problem. This task is accomplished by using certain mining method, algorithm or technique. It has been observed that the finer the pre-processing task is done of the raw data, the more useful and suitable information is possible to discover [4]. A schematic illustration of the proposed methodology is depicted in Figure 1.

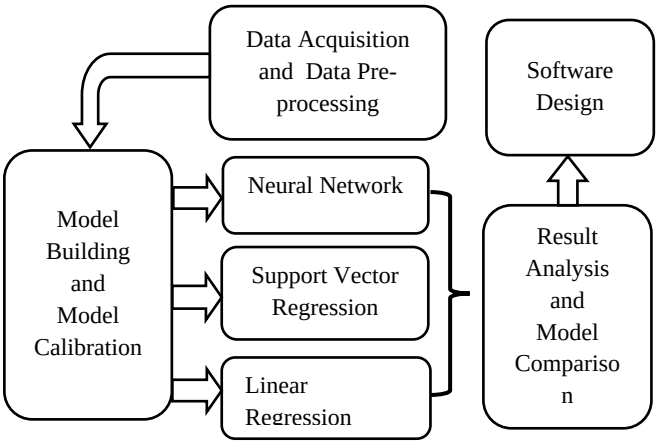

Figure 1: Proposed methodology

3.2 Methodology

3.2.1 Data Collection

The dataset used in this study was collected from Student Information System(SIS) of Hawassa University for the School of Computer Science. The dataset comprised 134 undergraduate degree students graduated from the university in the year 2015, 2016 and 2017 which consisted of 52(38.81%), 39(29.10%) and 43(32.09%) students respectively that managed to reach the final at semester eight. Those that failed along the way had been omitted from the study. The collected data was organized in Microsoft Excel sheet. Each student record had the following attributes:

Table 1. Student related attributes

No	Attributes	Description
1	Student ID	The student ID number
2	Sex	The sex of the student
3	Mobile No	The mobile no of the student
4	Section	The section into which the student assigned
5	Entry year	The year the student enrolled
6	Nationality	The nationality of the student
7	University Entrance Examination Result (UEER)	The result at the end of two years university preparatory completion examination at Grade 12
8	Course scores	The grade obtained by the student i.e. A, B, C, D, F
9	Semester GPA	The student GPA at the end of each semester
10	Final CGPA8	The student CGPA at semester eight (graduation)

3.2.2 Data Cleaning

The dataset collected from the university had some common mistakes such as inaccuracies, missing score for a course and inconsistent data. Therefore, to achieve the study goal attributes such as sex, mobile no, section, entry year, nationality, and UEER were insignificant for the study. In this manner, they were cleaned from the dataset.

3.2.3 Normalization

Normalization is a transformation of data in order to meet the input requirements of various data mining algorithms. In this paper for each student, forty-two (42) data points were collected including the final CGPA8 at graduation (Y) and the values of forty-one predictor variables (from X1 to X41) and a total of 134 students, 134 × 42 = 5,628 data points were collected. The collected data (Y, X1, X2, X3, ……., X41) were initially in different scales of measurements: X1 – X35 varied from A to F (letter grades), X36 – X41 varied from 0.00 to 4.00, and Y varied from 0.00 to 4.00. Before using them to establish prediction model, the collected raw data must be pre-processed, which is described in the following paragraphs.

First, all letter grades from X1 – X35, were converted into the corresponding numerical values (success coefficients), so regression models could be developed. As a result, the letter grades were converted to success coefficients as shown in Table 2.

Table 2. Course score conversion table

Letter grade	Success Coefficients/ Numerical value
A+	4.0
A	4.0
A-	3.75
B+	3.5
B	3.0
B-	2.75
C+	2.50
C	2.00
C-	1.75
D	1.0
F	0.0

Then, the numerical values of all data were normalized, so each data varied within the same range from 0 to 1, as shown in Table 3. The purpose of data normalization was to avoid the cases in which one variable received a high or low weight due to its initial low or large scale of measurements. This thesis adopts the min-max type of normalization which performs a linear transformation on the original data and has provides the highest accuracy.

Table 3. Normalization table

Predictor variables	Meaning	Initial value	Normalized value
X1 – X35	Courses taken by the student	Letter grade A+, A, A-, B+, B, etc.	0.00-1.00
X36 – X41	Semester GPA	0.00-4.00(numeric value)	0.00-1.00

3.2.4 Tools Used

To apply the data mining algorithms and pre-process the datasets, this paper used WEKA toolkit, [14] a widely used software for data mining that was developed at the University of Waikato in New Zealand. This toolkit provides a wide range of different data mining algorithms implemented in JAVA. It has been widely used in educational data mining researchers and for teaching purposes. In addition to WEKA, SPSS was also used for Pearson Correlation Analysis, and Microsoft Visual Studio 2015 with C# platform was used for implementing the application.

3.3 Evaluation Metrics

To validate the prediction models, the 10-fold cross-validation was used. The training set was randomly divided into 10 parts, nine of which were for training and the rest for testing. The process was repeated 10 times and then the accuracy of the model was computed. The study used one target variable, overall final CGPA8 where CGPA is a numeric variable ranging from 0.00-4.00.

To compare prediction methods, the Root Means Square Error (RMSE) is used. The RMSE is the square root of the average of the total squared error between the predicted and target values as in the following equation:

$$RMSE = \left(\frac{\sum (\hat{Y}_i - Y_i)^2}{n} \right)^{1/2}$$

where $\hat{Y}_i$ and Y_i are the predicted and targeted values, and n is the total number of records. small RMSE values give an indication of good prediction of the target values. Generally, if there is no significant difference between the compared models, the simpler and easier model to interpret is preferred.

3.4 Feature Selection

To proceed with the experiments, feature subset selection is performed. Feature selection is the process of removing features from the dataset that are irrelevant with respect to the task that is to be performed. Feature selection can be extremely useful in reducing the dimensionality of the data to be processed by the learning algorithm, reducing execution time and improving predictive accuracy.

The following three combinations of predictors were considered to do the experiment with different input features and the same output(target) variable.

Scenario #1: The students' university course scores from the first 2 years (i.e., scores of 23 courses) were used for predicting final CGPA8.

Scenario #2: The students' university course scores from the first 3 years (i.e., scores of 35 courses) were used for predicting final CGPA8.

Scenario #3: The students' Semester GPA at the end of each semester from the first 3 years courses were used for predicting final CGPA8.

4. EXPERIMENTAL RESULTS

4.1 Experiment 1

In this experiment, the performance of NN, SVR, and LR methods were tested for the first scenario (Scenario #1) mentioned in section 3.4.

Table 4. Prediction result for Scenario #1

Prediction Methods	R (Correlation coefficient)	RMSE	Time(sec)
NN	0.9089	0.1900	0.78
SVR	0.9305	0.1608	0.03
LR	0.9239	0.1675	0.05

As the results indicate, all the three prediction methods performed reasonably well in predicting the student final CGPA8. Among the three prediction methods, SVR method produced the most accurate prediction results, in which 0.9305 correlation coefficient(R), and 0.1608 RMSE values were obtained with the 10-fold cross-validation test. The second most accurate result was obtained with the LR method. The recorded correlation coefficient(R) and RMSE values were 0.9239, and 0.1675 respectively.

The least accurate result was obtained by the NN method, in which the lowest correlation coefficient (R) 0.9089 and highest RMSE 0.19 values were recorded. Other than these quantitative performance evaluation results, prediction results were qualitatively evaluated by plotting both the predicted and targeted as shown in the Figure 2-4.

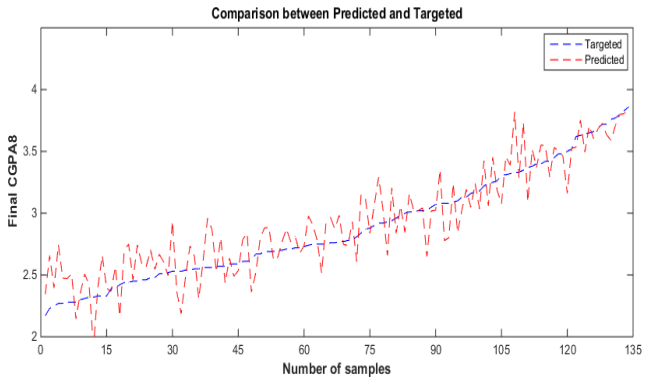

Figure 2: Prediction result of the NN method for the first scenario

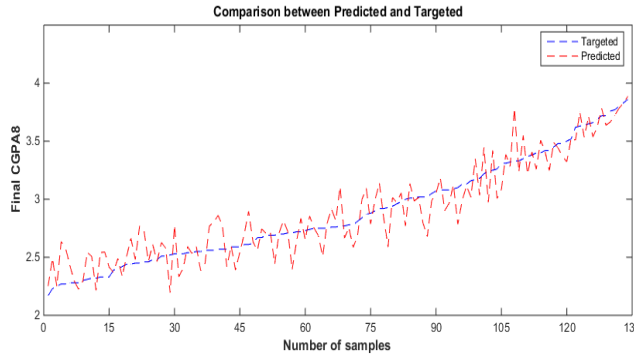

Figure 3: Prediction result of SVR method for the first scenario

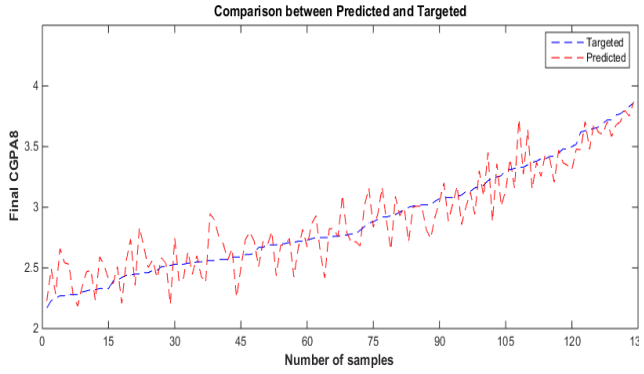

Figure 4: Prediction result of LR method for the first scenario

From Figure 2,3 and 4, it can be seen that, at lower CGPA8, the predicted is higher than the actual or targeted. At higher CGPA8, the predicted is lower than the actual CGPA8. As for middle CGPA8 values, there is no basic pattern or trend between predicted and targeted CGPA8.

4.2 Experiment 2

The experiments were repeated for the second scenario (Scenario #2). The results of the three prediction methods for the second scenario is presented in Table 5.

Table 5. Prediction result for Scenario #2

Prediction Methods	R (Correlation coefficient)	RMSE	Time(sec)
NN	0.9511	0.146	1.45
SVR	0.9742	0.0992	0.07
LR	0.9758	0.0954	0.02

The results indicate, all the three prediction methods performed reasonably well in predicting student final CGPA8. Among the three modeling methods, LR method produced the most accurate prediction results with 0.9758 correlation coefficient(R), and 0.0954 RMSE values were obtained with 10-fold cross-validation test. The second most accurate result was obtained by the SVR method with a correlation coefficient(R), and RMSE values of 0.9742, and 0.0992 respectively.

Again, the least accurate result was obtained by the NN method, for which the lowest correlation coefficient(R) 0.9511 and highest RMSE 0.146 values were recorded. Other than these quantitative performance evaluation results, the prediction results were qualitatively evaluated by plotting both the predicted and targeted, as shown in Figure 5–7.

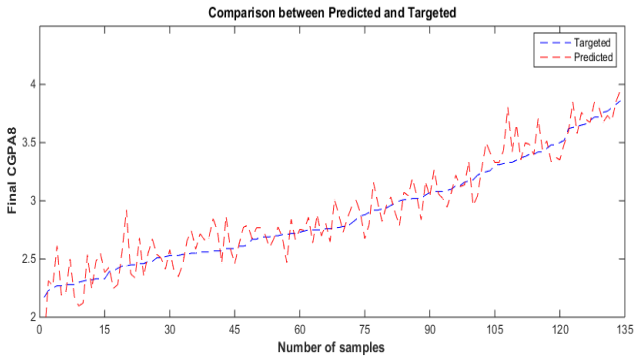

Figure 5: Prediction result of the NN method for the second scenario

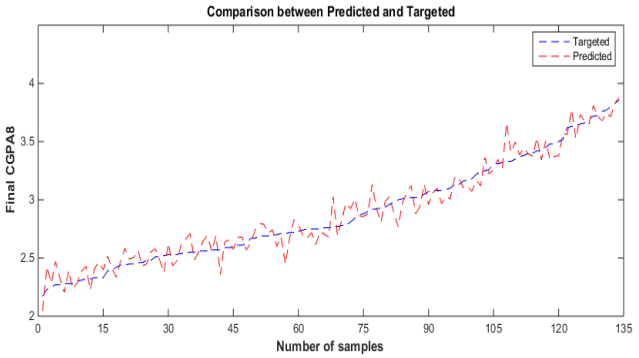

Figure 6: Prediction result of SVR method for the second scenario

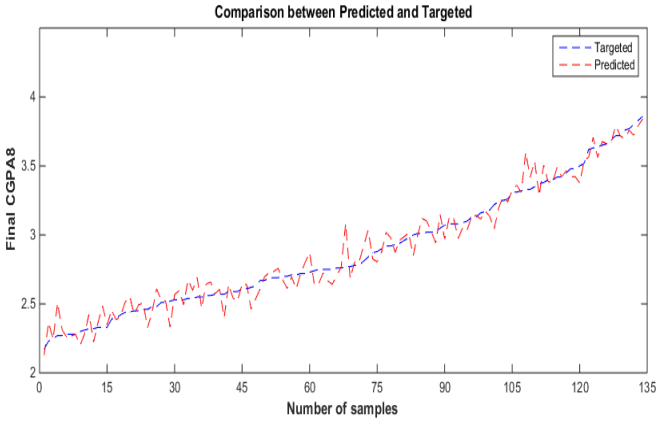

Figure 7: Prediction result of LR method for the second scenario

From Figure 5-7, it can be seen that there is no basic pattern or trend between predicted and targeted CGPA8 or there is no strong correlation between them except for Figure 7 where there was a slight variation for some data samples with high CGPA8 was predicted lower.

4.3 Experiment 3

In this experiment, the performance of the three prediction methods (NN, SVR, and LR) in predicting final CGPA8 was evaluated. The third scenario(Scenario #3) was considered to build the predictive models.

Table 6. Prediction result for Scenario #3

Prediction Methods	R (Correlation coefficient)	RMSE	Time(sec)
NN	0.9763	0.1000	0.08
SVR	0.9805	0.0862	0.02
LR	0.9805	0.0857	0.00

As can be seen from the above table, the third scenario took 0.08, 0.02 and 0 seconds to build the models for NN, SVR and LR methods respectively. All the three prediction methods performed reasonably well in predicting student final CGPA8. Among the three methods, LR and SVR produced equal prediction result of the correlation coefficient(R) 0.9805, and RMSE values of 0.0857 and 0.0862 respectively. Relatively, the least accurate result was attained by the NN method with a correlation coefficient(R) 0.9763, and RMSE value of 0.1. The figures shown below indicates the prediction results evaluated by plotting the predicted and targeted.

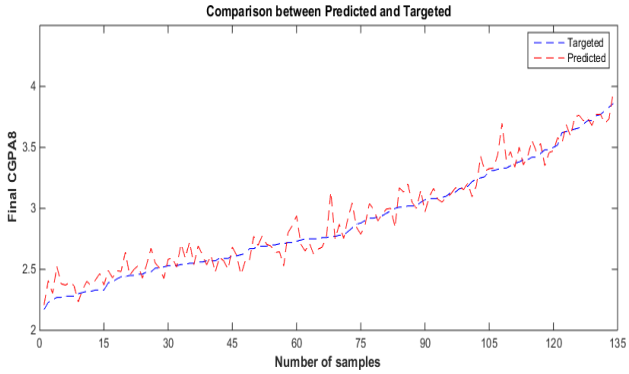

Figure 8: Prediction result of the NN method for the third scenario

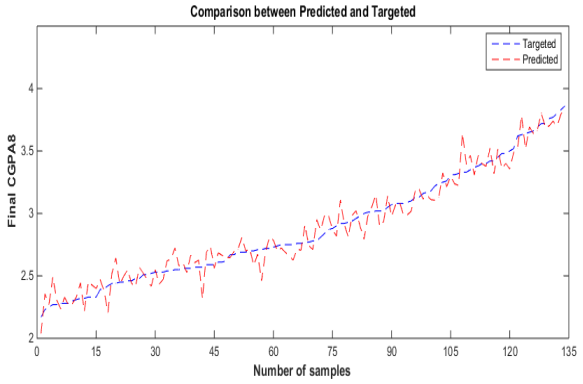

Figure 9: Prediction result of SVR method for the third scenario

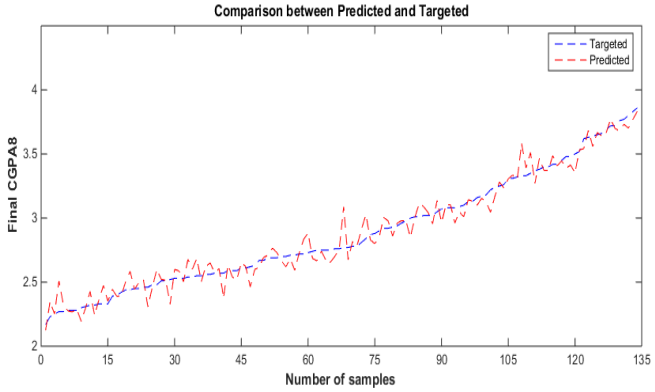

Figure 10: Prediction result of LR method for the third scenario

From Figure 8, we observed that there was a slight variation between predicted and targeted CGPA8. It can be seen that at lower CGPA8, the predicted is higher than the targeted CGPA8, while high CGPA8 predicted lower. As for Figure 9 and 10, it can be seen that at lower CGPA8, there is no basic pattern or trend on the predicted and targeted while those with high CGPA8 was predicted lower.

4.4 Experimental Analysis

Following the experiments conducted in this study, the next step was comparing the models and the best-trained model was then used to design a desktop application called Students' Performance Prediction System (SPPS).

As the experimental result indicate, SVR has the shortest time to build the model, while LR has the second shortest time and NN took the longest time for the first scenario. Regarding the second and third scenario, LR has the shortest time to build the models, while SVR has the second shortest time and NN took the longest time of all the three prediction methods. Regarding the correlation coefficient(R) and RMSE the result shows that LR was the best, SVR was the second best while NN was the least accurate of the three. This indicates that LR method outperforms the other two prediction methods. From this analysis, it can be concluded that models with the Linear Regression (LR) technique are more accurate and efficient. Therefore, this study proposes to use Linear Regression (LR) and Support Vector Regression (SVR) as the final prediction methods for developing Students' Performance Prediction System (SPPS).

5. SYSTEM DEVELOPMENT

This section discusses the designing and implementation of an application for predicting students' final CGPA8 using the models generated by SVR and LR prediction methods.

5.1 System Design

5.1.1 Use case Diagram

The use case diagram in Figure 11 depicts the possible interaction between the user and SPPS interface. Arrows that moves from the user to the system indicates user making prediction request by supplying the necessary information to the system while the one that comes out from the system indicates the system's response to the request made. The arrows that move from the administrator into the system indicates that the administrator provides the necessary backend assistant. The administrator has been added because the system needs someone to manage and service the system.

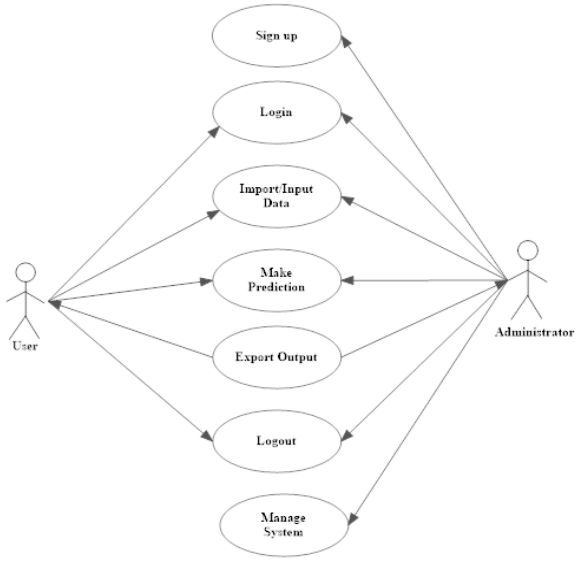

Figure 11: Use case diagram

5.1.2 Sequence Diagram

A sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. The sequence diagram for the system is given in Figure 12. The diagram has five main objects which are shown on the top of the diagram in a rectangle box with their class name. The five objects are Student, Login, Master, Predict, and Result.

The communication of information between two objects is represented by an arrow and message on that arrow. The vertical lines show the lifespan of the objects. According to the sequence diagram a user first login to the system and open the master page. After Master Page displayed, a user can click on Forms menu to get the Predict submenu to load, and select two more options to view the result.

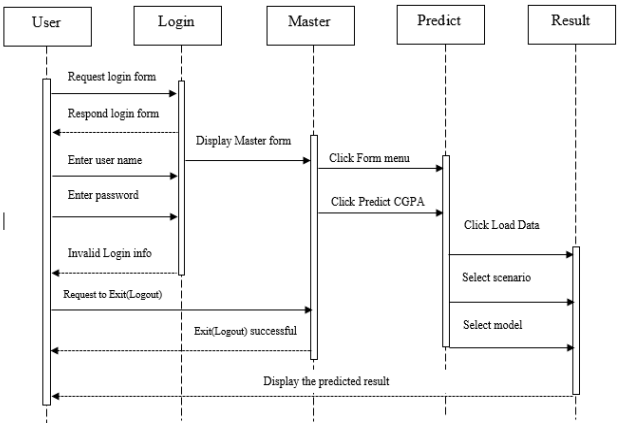

Figure 12: Sequence diagram

5.2 System User Interfaces

This section describes the user interfaces available on the system. The user interfaces of the system are designed for ease-of-use.

5.2.1 Login Page

The login page serves as an introductory page to the user. It gives access to the prediction system. The username and the password are security check to grant access to the system as shown in Figure 13. After successful login, the system directs the user to Master Page.

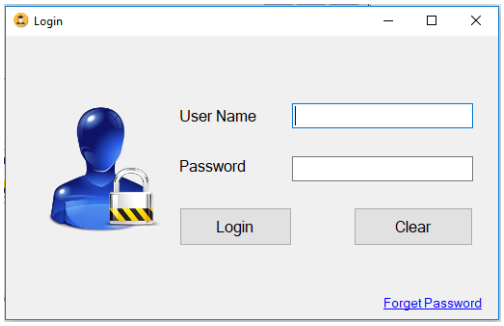

Figure 13: Login Page

5.2.2 Master Page

After successful login, the following screen is displayed for the user. A user can exit the system by selecting File→Exit menu

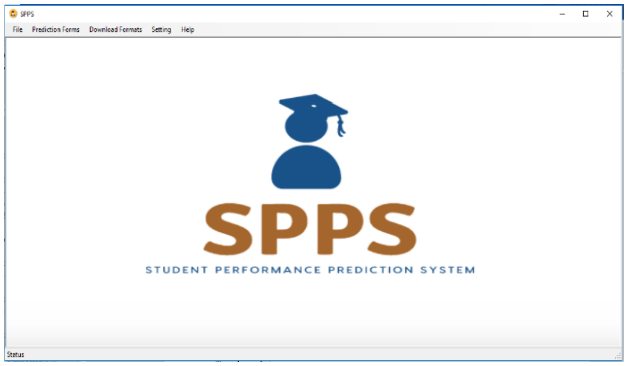

Figure 14: Master page

5.2.3 Prediction page

The user gets to this screen by selecting Forms → Predict students CGPA menu. This screen allows the user to make a prediction of CGPA and save the result to Microsoft Excel file.

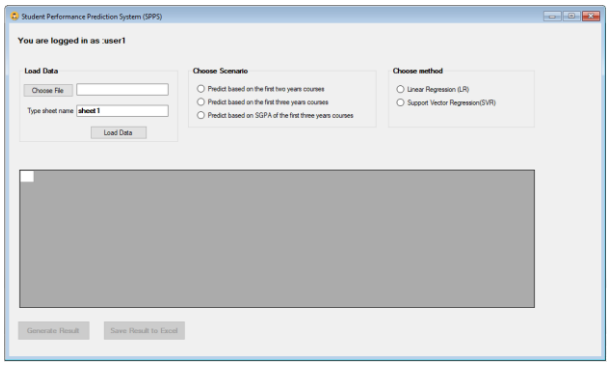

Figure 15: Prediction page

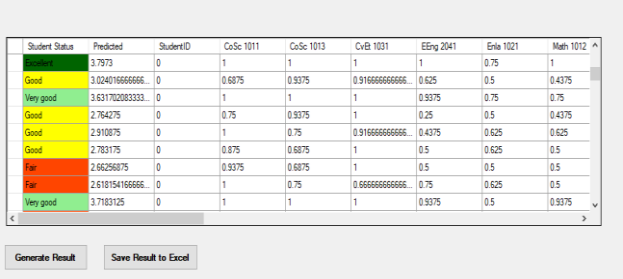

Figure 16: Sample output

6. CONCLUSION

The main purpose of this study was to develop a model for predicting students final CGPA8 and design an application based on the predictive models. The real dataset employed in this study was gathered from Hawassa University School of Computer Science graduates of 2015, 2016, and 2017.

In this study, experiments were done with three prediction methods-NN, SVR and LR to estimate student final CGPA based on three scenarios. The performance of the models was measured using the coefficient of Correlation (R) and that of Root Means Square Error (RMSE).

The first scenario was designed to predict final CGPA8 of students according to their university course scores completed during their 2 years of study. As the experimental result showed SVR method is efficient at minimizing the root mean square error between predicted and targeted. Besides, with SVR method student final CGPA8 prediction is possible at correlation coefficient(R)value equals to 0.9305.

In the second scenario, the university course scores during their 3 years of their coursework were used as input. As the cross-validation result indicates, LR was more efficient at minimizing the root mean square error between the predicted and targeted values and capable to predict the final CGPA8 at a correlation coefficient(R) value equals to 0.9758.

In the third scenario, the students' Semester GPA at the end of each semester during their 3 years of study were considered. As the result indicate, again LR method showed a slight improvement in minimizing the root mean square error between predicted and targeted. In addition, with LR method student final CGPA8 prediction is possible at a correlation coefficient value equals to 0.9805, thereby, further increasing the correlation coefficient(R) value by 0.0047.

Overall, the least accurate prediction result for all scenarios was obtained by the NN method.

The study also shows that it is possible to predict the student graduation performance, which is measured by CGPA using only scores of the first, second and third-year courses, no socio-economic or demographic features. Therefore, if a reasonable prediction can be reached without socio-economic data, it makes the implementation of student performance support system in a university easier.

From the study, it could be concluded that data mining techniques can be used efficiently for modeling and predict students' final CGPA8 in higher educational institutions and the models can also be used to design a predictive tool.

6.1 Future Works

There are several limitations of this study. The datasets were from one university and one cohort in the School of Computer Science, further research could include datasets from other programs and other institutions to rule out program or university bias.

Besides, educational research shows that some socio-economic, psychological factors, such as learning style, self-efficacy, motivation and interest, and teaching and learning environment, also play a role in student learning and thus affect student achievement. Therefore, future studies should include those above-mentioned variables in the models so as to increase the prediction accuracy [15].

7. REFERENCES

[1] Jiawei H, Kamber M. Data mining: Concepts and techniques, (the morgan kaufmann series in data management systems), vol. 2.: Morgan Kaufmann.

[2] Al-Razgan M, Al-Khalifa AS, Al-Khalifa HS. Educational data mining: A systematic review of the published literature 2006-2013.: Springer, 2014. p. 711-9.

[3] Asif R, Merceron A, Pathan MK. Predicting student academic performance at degree level: a case study. International Journal of Intelligent Systems and Applications. 2014;7：49.

[4] Asif R, Merceron A, Pathan MK. Predicting student academic performance at degree level: a case study. International Journal of Intelligent Systems and Applications. 2014;7：49.

[5] J. MM, S. H. Prediction of students' academic performance: Adapt a methodology of predictive modeling for a small sample size. 2014 IEEE Frontiers in Education Conference (FIE) Proceedings2014. p. 1-3.

[6] Baradwaj BK, Pal S. Mining educational data to analyze students' performance. arXiv preprint arXiv:1201.3417. 2012.

[7] Goga M, Kuyoro S, Goga N. A recommender for improving the student academic performance. Procedia-Social and Behavioral Sciences. 2015;180：1481-8.

[8] Mishra T, Kumar D, Gupta S. Mining students' data for performance prediction. 2014. p. 255-62.

[9] Nandeshwar A, Chaudhari S. Enrollment prediction models using data mining. Retrieved January. 2009;10：2010.

[10] Garcia-Saiz D, Zorrilla M. Comparing classification methods for predicting distance students' performance. 2011. p. 26-32.

[11] Y. YC, S. MT, C. SCN. Determinants of student performance in advanced programming course. 2012 International Conference for Internet Technology and Secured Transactions2012. p. 304-7.

[12] Golding P, Donaldson O. Predicting academic performance.: IEEE, 2006. p. 21-6.

[13] Altaher A, BaRukab O. Prediction of Student's Academic Performance Based on Adaptive Neuro-Fuzzy Inference. International Journal of Computer Science and Network Security (IJCSNS). 2017;17：165.

[14] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM SIGKDD explorations newsletter. 2009;11：10-8.

[15] Lin JJ, Imbrie PK, Reid KJ. Student retention modelling: An evaluation of different methods and their impact on prediction results. Research in Engineering Education Symposium. 20091-6.