Factors Affecting Secondary School Students' Academic Performance, Kenya

Obadiah Matolo Musau[1], Kelvin Omieno[2], Raphael Angulu[3]

[1]*School of Computing and Informatics, Technical University of Mombasa, Kenya*

[2]*Department of Information Technology and Informatics, Kaimosi Friends University College, Kenya*

[3]*Department of Computer Science, Masinde Muliro University of Science and Technology, Kenya*

Abstract **-Student's academic performance in learning environments is linked to several factors that include student demographics, educational background and other environmental variables. The purpose of this study was to identify the most influential factors that affect secondary school students' KCSE performance in Kenya. The research is based on reviewed existing literature on students' academic performance andsecondary school student'sdata collectedin Kenya through use of questionnaires.**

Keywords **-Academic Performance, Feature Selection, Kenya Secondary School Examination**

I. INTRODUCTION

The need for quality monitoring of student academic performance in any learning process can not be underestimated. However, this comes with unique challenges such as how todetermine the parameters that influence student academic performance. In order to effectively address the rising concerns in student performance such as student dropout prior to completing the school cycle, high rate of class repetition and poor academic performance(UNICEF, 2009), it is important to know the crucial factors that affect students' success in their educational environments. This study aimed to contribute to this issue by finding out the factors responsible for success and failure of students in secondary schools in Kenya. A research was conducted between January and April 2019 to collect information on factors affecting students' academic performance.

II. LITERATURE REVIEW

Most of the literature reviewed on student academic performance support the assertion that success in academics is dependent on several factor which include a mixture of student demographic features, educational background and environmental features.Paulo & Silva (2008) explainedthat secondary school students performance is affected by very diverse factors that include student previous academic achievements, students' demographic, social and school related attributes such as student's age,student absences, alcohol consumption,parent's job and mother's education.

Oladokun, Adebanjo, & Charles-Owaba (2008) carried out a study on student academic performance and found that parental background, gender,ordinary level subjects' scores, subject's combination, matriculation examination, scores,type of school, location of schooland age on admission have high effect on student academic performance. According toOsmanbegović & Suljić (2012),gender, distance, GPA, scholarship, learning materials and previous grades influence students' academic performance. Khasanah & Harwati (2017) identified that attributes:gender, origin, father education, father occupation, mother education, mother occupation, senior high school type, senior high school department, senior high school final grade, attendance, GPA and drop out were the most influential student attributes in determining academic performance.

Khan, Hayat, & Daud (2015) identified student's previous grades as the most influential component in determining student academic performance. It was also noted that parents' occupation played a major role in students' academic performance than the type of school. Similar study by Asif, Merceron, & Pathan (2014) explained that student performance is dependent onprevious academic achievements such as pre-admission marks, first year courses marks and second year courses marks.Sundar (2013) predicted academic performance of students based on previous performance, assignment marks, attendance, internal marks, seminar, and co-curricular activities.Table I shows a comprehensive literature review of recent studies on students' academic performance.

TABLE I

FACTORS USED TO PREDICT STUDENT ACADEMIC PERFORMANCE

Source	Factors
Livieris, Drakopoulou, Tampakas, Mikropoulos, & Pintelas (2018)	Secondary stage type, Oral grade of the first test, second test and final examination of the first and second semester, Final grade of the first and second semester and Grade in the final examinations
Gadhavi & Chirag (2017)	average of unit test and sessional examination marks
Khasanah & Harwati (2017)	gender, origin, father education, father occupation, mother education, mother occupation, senior high school type, senior high school department, senior high school final grade, attendance and GPA
Sharma & Santosh (2017)	roll number, name, assignm1, assignm2, midsem1, midsem2 and final performance of the students in that semester
Kaur & Singh (2016)	gender, hometown, family income, previous semester grade, attendance, medium(language) and senior secondary grade, seminar performance and sports.
Khan, Hayat, & Daud (2015)	Student marks in SSC-I, final grade in SSC-II and number of students

Shahiri, Mohamed, & Husain (2015)	Internal assessments, psychometric factors, external assessment, CGPA, student demographic, high school background, scholarship, social network interaction, extra-curricular activities and soft skills
Agrawal & Mavani, March (2015)	Student's grade in secondary education, living location and medium of teaching.
Asif, Merceron, & Pathan (2014)	4th year grade, HSC examination total marks, HSC examination mathematics marks, marks for units: MPC, CT-153, CT-157, CT-158, HS-205/206, MS-121, CS-251, CS-252, CT-251, CT-254, CT-255, CT-257, EL-238 and HS-207
Sundar (2013)	student id, name, quota in which student joins, previous semester performance, performance in internal exam, performance in seminars, assignment, attendance, co-curriculum activities and end of semester marks
Osmanbegović & Suljić (2012)	Gender, family, distance, high school, GPA, entrance exam, scholarships, time, materials, internet, grade importance and earnings
Kabakchieva (2012)	Gender, birth year, birth place, living place and country, type of previous education, profile and place of previous education, total score from previous education, university admittance exam and achieved score, total university score at the end of the first year and number of failures
Lin (2012)	gender, state, citizenship, academic major, ethnic group, age, student aid, family contribution, financial need, loan received, awarded scholarship and cal grant receiver
Bhardwaj & Pal (2011)	sex, student category, medium of teaching, student food habit, student other habit, living location, hostel, family size, family status, family income, students grade in senior secondary education, student's college type, father's qualification, mother's qualification, father's occupation, mother's occupation and grade obtained in BCA
Oladokun, Adebanjo, & Charles-Owaba (2008)	UME score, O'level results, further math, age at entry, time before admission, parents education, zone of secondary school attended, type of secondary school, location of school and gender.
Paulo & Silva (2008)	Sex, age, school, address, parents cohabitation status, mothers education, mothers job, fathers education, fathers job, family size, guardian, family relationship, reason for choice of school, travel time, study time, failures, school support, family support, activities, extra paid classes, internet, nursery, higher education interest, romantic, free time, going out with friends, alcohol consumption, health status, absences, first, second and third period grades.

However, prediction of student's academic performance in the developing countries has not been sufficiently investigated. The objective of this study was to find out the most influential factors in determining students' academic performance.Theworkwas based on a study carried to collect information from secondary school students in Kenya.

III. METHODOLOGY

The sample data was collected between January 2019 and April 2019from 1720 students. It consisted of a mixture of students that studied from both public and private secondary schools.The data was collected through a questionnaire administered randomly to the participants by research assistants. The questionnaire consisted of a total of 35 questions printed in 4 standard A4 size sheets. The questions were further divided into 6 sections; section 1 consisted of general questions (such as type of institution, date etc.), section 2 was on students' demographic attributes, section 3 collected data on students' family information, section 4 collected data on co-curriculum information, section 5 collected data on previous students' academic performance and section 6 collected data on school demographic features.All the experiments were carried out on WEKA machine learning environment. Feature selection techniques were used to rank the attributes by evaluating the usefulness of each attribute in predicting students' academic performance.

IV. DATA DESCRIPTION

The sample comprised of 1720 instances and 60 attributes. The class variable was Kenya secondary school examination (KCSE) grade which is the final academic performance of the student. The other variables served as independent variable. The attributes information and the associated domains are shown in Table II.

TABLE II STUDENT RELATED ATTRIBUTES

No	Attribute Name	Description	Domain
1	Gender	Gender	{female,male}
2	Age	Age	{ Below 14 yrs (1), 14-18 yrs(2), above 18yrs(3)}
3	Disability	Disability	{yes,no}
4	Religion	Religion	{muslim,christian,others}
5	LP	Lived with Parents	{yes,no}
6	WPC	Witnessed Parent Conflicts	{yes,no}
7	FS	Family Structure	{singleparent,nuclear,extended,step}
8	DF	Difficulties Paying Fees	{yes,no}
9	Sponsor	Sponsor	{parents,guardian,others}
10	PE	Parents Employment	{ both(1),one(2),none(3)}
11	FE	Father's Education	{none(1),primary education(2),secondary education(3),postsecondary(4),degree and above(5)}
12	ME	Mothers Education	{none(1),primary education(2),secondary education(3),postsecondary(4),degree and above(5)}
13	CA	Participated in Curriculum Activities	{yes,no}
14	CF	Frequency of Participation	{1,2,3,4,5}
15	NSF1	Number of subjects in Form 1	{7,8,9,10,11,12,13,14,15,16}
16	NSF2	Number of subjects in Form 2	{6,7,8,9,10,11,12,13,14,15}
17	NSF3	Number of subjects in Form 3	{4,5,6,7,8,9,10,11,12}
18	NSF4	Number of subjects in Form 4	{4,5,6,7,8,9,10,11,12}
19	Specialization	Specialization	{yes,no}
20	YS	Year of	{1,2,3,4}

		Specialization	
21	CS	Changed School	{yes,no}
22	RC	Repeated Class	{yes,no}
23	F1G	Form 1 Grade	{a,b,c,d,e}
24	F2G	Form 2 Grade	{a,b,c,d,e}
25	F3G	Form 3 Grade	{a,b,c,d,e}
26	MG	Mock Grade	{a,b,c,d,e}
27	KCSE	KCSE Grade	{a,b,c,d,e}
28	LS	Learning Styles Used	{ one(1),two(2),three(3)}
29	AS	Assessment Style	{ formal(1),informal(2),all(3)}
30	ELS	Effect of Learning Style	{very low(1),low(2),moderate(3),high(4),very high(5)}
31	EAS	Effect of Assessment Style	{very low(1),low(2),moderate(3),high(4),very high(5)}
32	Absences	Absences in months	{0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,18,24}
33	EC	Examination Challenges	{yes,no}
34	AD	Access to Drugs	{yes,no}
35	RM	Role Model	{yes,no}
36	EA	Effect of Absences	{very low(1),low(2),moderate(3),high(4),very high(5)}
37	ETA	Effect of Teacher Absenteeism	{very low(1),low(2),moderate(3),high(4),very high(5)}
38	EFS	Effect of Failure to Cover Syllabus	{very low(1),low(2),moderate(3),high(4),very high(5)}
39	ECA	Effect of Co-Curriculum Activities	{very low(1),low(2),moderate(3),high(4),very high(5)}
40	ED	Effect on Access to Drug	{very low(1),low(2),moderate(3),high(4),very high(5)}
41	EEC	Effect of Examination Challenges	{very low(1),low(2),moderate(3),high(4),very high(5)}
42	ERM	Effect of Role Model	{very low(1),low(2),moderate(3),high(4),very high(5)}
43	TS	Type of School	{national,extracounty,subcounty,county}
44	Residence	Residence	{ boarding(1),day(2),both(3)}
45	SC	School Composition	{girls,mixed,boys}
46	TL	Teaching Laboratory	{yes,no}
47	Library	Library	{yes,no}
48	CL	Computer Laboratory	{yes,no}
49	Electricity	Availability of Power/Electricity in school	{yes,no}
50	Internet	Access to Internet	{yes,no}
51	STL	Status of Teaching Laboratory	{worst(1),worse(2),bad(3),good(4),better(5),best(6)}
52	SL	Status of Library	{worst(1),worse(2),bad(3),good(4),better(5),best(6)}
53	SCL	Status of Computer Laboratory	{worst(1),worse(2),bad(3),good(4),better(5),best(6)}
54	SE	Status of Power/Electricity	{worst(1),worse(2),bad(3),good(4),better(5),best(6)}
55	SI	Status of Internet	{worst(1),worse(2),bad(3),good(4),better(5),best(6)}
56	ETL	Effect of Teaching Laboratory	{very low(1),low(2),moderate(3),high(4),very high(5)}
57	EL	Effect of Library	{very low(1),low(2),moderate(3),high(4),very high(5)}
58	ECL	Effect of Computer Laboratory	{very low(1),low(2),moderate(3),high(4),very high(5)}
59	EE	Effect of Power/Electricity	{very low(1),low(2),moderate(3),high(4),very high(5)}
60	EI	Effect of Internet	{very low(1),low(2),moderate(3),high(4),very high(5)}

V. FEATURE SELECTION

In order to get a better understanding of the significance of each attribute to the class variable, we analysed the impact of each attribute in relation to the dependent variable using feature selection techniques. Feature selection is a technique used to select the most relevant features (attributes) in predicting the class variable(Khasanah& Harwati, 2017; Ramaswami & Bhaskaran, 2009). We conducted experiments on WEKA using three commonly used feature selection techniques; Gain Ratio, Info Gain and One R-test. Feature selection in WEKA involves two major steps; selection of the attribute evaluator and selection of a search method. Attribute evaluator is a technique which evaluates each attribute in the dataset in the context of the class attribute, and the search method is the technique used to navigate different combinations of features in the dataset before settling on the chosen features. In order to get the final subset of the features that strongly correlate with the class attribute, we compared the results of each technique using the name of the attribute and the measure of goodness (merit) as the evaluation metrics. However, it was noted that each method has a unique way of accounting for the significance of the attributes. An average value of the values provided by each feature selection techniques used was therefore takenas a representative of the final value instead of selecting one technique or method over others. A summary of the results obtained using feature selection techniques are shown in Table III.The results showed MG as the most significant attribute in determining student academic performance.

TABLE III SUMMARY OF FEATURE SELECTION

N	Attribute	CBFS	IGBFS	OneR	Average
1	MG	0.342	0.30719	68.256	22.96839667
2	F3G	0.2841	0.23445	63.14	21.21951667
3	ME	0.1104	0.0988	62.791	21.00006667
4	FE	0.1074	0.09609	62.267	20.82349667
5	NSF2	0.0774	0.07421	61.628	20.59320333
6	AS	0.1341	0.04833	61.163	20.44847667
7	F2G	0.1032	0.10406	60.174	20.12708667
8	Religion	0.1188	0.0243	60.116	20.08636667
9	NSF1	0.1278	0.05705	60	20.06161667
10	DF	0.1904	0.04557	59.826	20.02065667
11	F1G	0.1763	0.11178	59.767	20.01836
12	Internet	0.1594	0.04024	59.826	20.00854667

13	EC	0.1612	0.03136	59.826	20.00618667
14	CL	0.1142	0.02659	59.826	19.98893
15	RM	0.1151	0.02072	59.826	19.98727333
16	STL	0.0727	0.05947	59.826	19.98605667
17	SL	0.0705	0.059	59.826	19.98516667
18	LS	0.1009	0.02841	59.826	19.98510333
19	Library	0.1001	0.0284	59.826	19.98483333
20	Residence	0.0989	0.0291	59.826	19.98466667
21	YS	0.0789	0.04651	59.826	19.98380333
22	NSF3	0.0394	0.0247	59.884	19.9827
23	SI	0.0636	0.0545	59.826	19.98136667
24	TS	0.0619	0.05238	59.826	19.98009333
25	CA	0.1001	0.01076	59.826	19.97895333
26	SC	0.0769	0.02933	59.826	19.97741
27	SCL	0.0486	0.05033	59.826	19.97497667
28	EAS	0.0599	0.03217	59.826	19.97269
29	Gender	0.0703	0.02125	59.826	19.97251667
30	Specialization	0.0731	0.01029	59.826	19.96979667
31	Age	0.0665	0.01321	59.826	19.96857
32	Sponsor	0.0698	0.00858	59.826	19.96812667
33	CF	0.0483	0.02899	59.826	19.96776333
34	SE	0.0411	0.03417	59.826	19.96709
35	EI	0.049	0.02199	59.826	19.96566333
36	ECL	0.0439	0.0227	59.826	19.9642
37	FS	0.0529	0.01056	59.826	19.96315333
38	TL	0.0496	0.01379	59.826	19.96313
39	ED	0.0473	0.01346	59.826	19.96225333
40	CS	0.0498	0.00392	59.826	19.95990667
41	EE	0.0377	0.01133	59.826	19.95834333
42	ETL	0.0281	0.01963	59.826	19.95791
43	Electricity	0.0405	0.00621	59.826	19.95757
44	EA	0.0318	0.01446	59.826	19.95742
45	EFS	0.0301	0.01187	59.826	19.95599
46	EL	0.0273	0.01303	59.826	19.95544333
47	ERM	0.0254	0.01344	59.826	19.95494667
48	ETA	0.0261	0.01243	59.826	19.95484333
49	ECA	0.0215	0.01027	59.826	19.95259
50	RC	0.0296	0.00163	59.826	19.95241
51	EEC	0.0144	0.01412	59.826	19.95150667
52	Disability	0.0237	0.00307	59.826	19.95092333
53	WPC	0.0234	0.00106	59.826	19.95015333
54	AD	0.0105	0.01188	59.826	19.94946
55	LP	0.0115	0.00156	59.826	19.94635333
56	Absences	0.1208	0.05252	59.593	19.92210667
57	ELS	0.0553	0.02482	59.651	19.91037333
58	NSF4	0.0527	0.02181	59.186	19.75350333
59	PE	0.1222	0.06065	58.372	19.51828333

VI. MODEL

In order to select the most influential attributes from the feature vector of 60 attributes, we performed several experiments to find out the best combination of attributes that predicted the dependent variable with high accuracy. Three machine learning algorithms were used: Naïve Bayes, J48 Decision Tree and Multilayer Perceptron (Khan, Hayat, & Daud, 2015; Baradwaj & Pal, 2011; Dey, 2016; Osmanbegović & Suljić , 2012; Agrawal & Mavani, March 2015).Selection of the optimal feature subset was done by successive modeling where training of the models started with an initial feature subset of the top ranked three features then proceeded successively adding a feature in each iteration until each model reached the optimal performance.

The best accuracy of 73.02% by J48 Decision Tree was achieved using the top 14 ranked attributes as the input variables. The results of the experiments are shown in Table IV.

TABLE IV

PERFORMANCE OF MODELS BASED ON THE SELECTED FEATURES

Evaluation Metric	Model		
	Naïve Bayes	**J48**	**Multilayer Perceptron**
Accuracy	71.45%	73.02%	71.63%
Correctly Classified Instances	1229	1256	1232
Incorrectly Classified Instances	491	464	488

VII. DISCUSSION

The aim of this study was to identify the factors that affect secondary school students' final performance in Kenya. The study found out that the Mock Examination as the most influential factor. This could possibly be because mock examinations are often used as an indicator of how well a student is prepared for the final examination. Mock grade is the test grade students score before sitting for the final KCSE examination in secondary school hence it is a significant factor for predicting KCSE examination grade. Other previous examination performances like form three grade and form two grade were listed to be influential, however, form one grade was not found to be very influential.

Another factor was the number of subjects taken by the students. The study found out that as number of subjects reduce due to specialization at higher classes, there is slight improvement in performance. Challenges experienced by candidates during examination periods were also identified as being influential in determining students' performance. These challenges include bereavement, sickness, etc. Mothers' education and fathers' education were found to be influential in students' academic performance. Religious background was also listed among the top factors affecting student performance. Financial challenges especially in fees payment and the presence of role model figure were other factor found to affect academic performance. School related factors that were found to affect students' academic performance included teaching laboratories, computer laboratories, library and type of residence, i.e., boarding or day school.

VII CONCLUSION

In this study we started by carrying out a systematic literature review of studies on students' academic performance in order to find out the factors that affect students' academic performance. A fact-finding study was conducted to find out the effect of various factors on the secondary school students academic performance in Kenya. Feature selection techniques were used to rank the attributes based on the usefulness of each feature in determining the final academic performance of the student in KCSE. Finally, experiments were conducted using machine learning techniques to determine the most influential factors. The results showed that mock examination, previous examination performances, number of subjects, examination related challenges, mothers' education, fathers' education, religion, financial challenges, teaching laboratories, computer laboratories, presence of role model, school library and type of residence were the most influential factors.

REFERENCES

[1] Agrawal, H., & Mavani, H. (March 2015). Student Performance Prediction using Machine Learning. International Journal of Engineering Research & Technology (IJERT) , 4(03), 111-113.

[2] Asif, R., Merceron, A., & Pathan, M. K. (December 2014). Predicting Student Academic Performance at Degree Level: A Case Study. International Journal of Intelligent Systems Technologies and Applications, 01, 49-61.

[3] Baradwaj, B. K., & Pal, S. (2011). Mining Educational Data to Analyze Students" Performance. International Journal of Advanced Computer Science and Applications, 2(6), 63-69.

[4] Dey, A. (2016). Machine Learning Algorithms: A Review. International Journal of Computer Science and Information Technologies, 7(3), 1174-1179.

[5] Khan, B., Hayat, M. S., & Daud, M. (2015). Final Grade Prediction of Secondary School Student using Decision Tree. International Journal of Computer Applications, 115(21), 32-36.

[6] Khasanah, A. U., & Harwati. (2017). A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques. IOP Conference Series: Materials Science and Engineering, 215(1).

[7] Oladokun, V. O., Adebanjo, A. T., & Charles-Owaba, O. E. (2008). Predicting students academic performance using artificial neural network: A case study of an engineering course. The Pacific Journal of Science and Technology, 8(1), 72-79.

[8] Osmanbegović, E., & Suljić , M. (2012). Data mining approach for predicting student performance. Economic Review: Journal of Economics and Business, 10(1), 3-12.

[9] Paulo, C., & Silva, A. (2008). Using data mining to predict secondary school student performance. 5-12.

[10] Ramaswami, M., & Bhaskaran, R. (2009). A Study on Feature Selection Techniques in Educational Data Mining. Journal of Computing, 1(1), 7-11.

[11] Sundar, P. P. (2013). A Comparative Study for Predicting Student's Academic Performance Using Bayesian Network Classifiers. IOSR Journal of Engineering (IOSRJEN) e-ISSN, 2250-3021.

[12] UNICEF, W. (2009). Abolishing School Fees in Africa: Lessons from Ethiopia, Ghana, Kenya, Malawi and Mozambique. Development Practice in Education.