International Journal of Computer Applications Technology and Research
Volume 13–Issue 05, 35 - 41, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1305.1008

An Automated System for Ranking of Scholarship Applicants on the Basis of Brightness and Poverty Levels

[1] Ekitoe, Paul, [2]Otieno, Calvins
[1,2] Department of Computer Science, Maseno University

Abstract: Access to education is a basic right. Available resources that could aid access to education should be distributed fairly and objectively to guarantee that all those that deserve have a fair share of access. The process of selecting and awarding scholarship where the population outstrips the available resources calls for a transparent, standardized selection criteria that can be executed quickly. The study sought to develop a system that ranks scholarship applicants based on two parameters; brightness and poverty levels. The study identified parameters for generating brightness and poverty level indices. The identification of the parameters was done through exploratory approach as well as using secondary data while the development and testing of the system used experimental approach. The outcome of the system ranking was compared to the actually ranking and award made in the period of successive five years. The outcome obtained showed significant variation in ranking between the manual and automated rankings. The statistical significance of the difference as confirmed using Chi-Square and Fisher's exact tests. This outcome tends to suggest that automated approach presents a more objective way of ranking where all applicants are evaluated using the same scale of reference. The system thus can be viewed as a tool with potential to enhance objectivity.

Keywords: Ranking, Scholarship, Brightness, Poverty Level, Ranking Algorithms

1.0 INTRODUCTION

Education is considered a basic need. A growing trend in developing countries, where an increasing number of people, depend on financial assistance to pursue education has been observed. The Cost sharing approach [1] as opposed to free education makes access to education expensive, disadvantaging the less privileged.

The challenges associated with financing education, and the universal effort to ensure access to education has birthed schemes and strategies that aim at making the cost-sharing element successful. These schemes include the award of bursaries, scholarship, loans and grants. In Kenya for example, the effort to assist the poor in access education include the establishment of the Higher Education Loans Board (HELB), a government agency that evaluates application and awards education sponsorship and loans based on level of need [2]. In addition, the National Government Constituency Development Fund, established in 2003 through an act of parliament has played a significant role financing education of needy students through award of scholarships and fee bursaries. It has been argued [3][4]that these initiatives have played a critical in providing equal opportunities for the poor to access education and thus lowering the dropout rate.

In resource constrained localities, a fair approach to distribution of available resources is needed. Ranking, which provides a list of individuals and their levels of need aids in objective and fair distribution of the limited resources. Ranking of scholarship applicants based on either merit or need, and using manual processes that requires 100% human intervention is complex and time consuming. The complexity arise especially where there an overwhelming number of applicants and ranking parameters that contain sub-attributes with complex set of intra and inter-relationship between them. In such manual processes, human bias thatleads to loss of objectivity in ranking cannot be ruled out. In cases where the process is automated, the systems, models or algorithms used are highly tailored and thus leave no room for possible adoption of such systems, models or algorithms for ranking applicants at different levels e.g. primary, secondary and university levels. In other cases, the systems are designed to predict level of need rather than rank applicants and thus excluded the critical component of brightness or academic performance hence limiting their ability to compute and give a true picture of level of the applicant's qualification. It has been observed [5] thatthere were no common, internationally accepted principles, tools and templates of administration and implementation of bursaries or scholarships.

In Kenyan scenario [2], is has been reported that even with the government efforts to reduce such cases of financial difficulties, there were serious concerns for the disadvantaged; identification, ranking and award of the limited financial aid was not done in a fair and objective manner. There were numerous cases where undeserving applicants were awarded while those deserving missed out on the allocation. Further, it has been argued [6][7]there are several factors that contribute to lack of objectivity and fairness in ranking of applicants; the lack of a clear way to determine who is needy, the weak administrative mechanisms and the gender bias that affects girl child.

Analysis of the Kenyan scenario, by different studies pointed out the existence of the following challenges:

- *Misallocation of scholarship funds, double awards to one student in two schools, awarded to "ghost" students, as well as excessive patronage by members of parliament who influence the allocation of the funds [8]*
- *Political influence where undue pressure is exerted on scholarship committee to allocate bursaries to undeserving relatives, thereby denying the genuinely needy students access to the justified award[9]*
- *The methods used to identify and rank the applicants faced multiple challenges that often lead to unfair award of bursaries or scholarship[10]*

Without a streamlined, standardized and objective process, challenges that include skewed outcomes and unfair wards in the ranking would persist.

International Journal of Computer Applications Technology and Research
Volume 13–Issue 05, 35 - 41, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1305.1008

2.0 REVIEW OF EXISTING RANKING METHODS

Automation means employing technology to a process in order to realize efficiency, objectivity and standardization. Gradually, there is a growing interest in the process of automation where technology has been used to replace the manual processes [11]. The goal has been to realize fairness and objectivity in ranking of applicants.

2.1 Manual Ranking

The manual process is the traditional approach used to identify and rank scholarship applicants. The process requires human intervention at all stages and the credibility of the outcome is largely dependent on the expertise and integrity of those charged with the responsibility of overseeing the process. It was observed[10] that although the approach has a level of flexibility in terms of the ability to interrogate applicants to establish certain pieces of pertinent information, it is time consuming, requires a huge amount of resources, it is prone to manipulation and bias and possibility of human errors makes the process unsuitable.

2.2 Regency Frequency and Monetary (RFM) Model and Clustering Algorithm

Regency Frequency and Monetary (RFM) model and clustering algorithm was used [12] to build a framework for identifying needy students and determine their loan valuation. RFM predicts the trends of a particular customer using three key pieces of information about customer's past behavior i.e. Recency; how long ago the customer made the last purchase, Frequency; the number of purchases made by the customer and Monetary; the amount spent by the customer. The study was based on a university in China where students get food supplies from a canteen which provides them with a smartcard, and the canteen database stores records of their transactions. The RFM model used information from the canteen database to identify needy students whereby the three attributes were calculated as:

- R (Recency) as ratio=average cost school fees,
- Frequency= the number of monthly consumptions,
- Monetary= family Monthly consumption

This model has limitations. Although past behavior on spending can, to some degree indicate the financial need, without considering the underlying parameters, the outcomes may be wrong. The model assumes that the applicants is already in the system and thus does not consider a first time applicant. In such a case, the model tends to be biased as it relies on the spending rates of students to determine whether they are qualified applicants.

2.3 Multi-Modal Multi-Label Approach

A Multi-modal Multi-label approach was proposed [13] to determine types and amount of grants to be given to students in universities. This approach used student's information collected through multi-modal channels; which include their behavior of internet usage, campus consumption, and daily trajectory together with their enrollment information that includes personal and family information. This model has limitations. The rank operator that is deemed to regularize the prediction results can be biased at certain level. The rank operator has to regularize the information to ensure modal consistency, and as such certain information may not be true in relation to the data collected. In this model, some of the elements used to rank e.g. online behavior may not apply to all cases, especially in remote villages in Kenya.

2.4 C4.5 Algorithm Tree

The C4.5 algorithm tree starts with training of the data in the single root nodes that is partitioned into nodes based on the splitting criterion. When dealing with the C4.5 algorithm, the default splitting criterion adopted uses the information gaining ratio. There are variables [14] in which the nodes are partitioned via the adoption of C4.5 algorithm. When the splitting is performed through the use of quantitative independent variable, there are partitions which are known as the threshold and could have other subsequent child nodes. In the C4.5 algorithm, there is also the utilization of the tree growing approach which mainly focuses on finding the best local choice. Therefore, the method is considered to utilize the ratios of information gained to help in splitting criterion for he qualified and the population that is not qualified to receive the financial support. It is clear that the model has a limitation in that the algorithm only relies upon the previous messages to predict an outcome. Therefore, the model is not appropriate for ranking the scholarship applicants because there is a likelihood that the previous students who had applied will qualify for the scholarship award and thus biased against new applicants who may be in greater need of such financial support compared to individuals who had applied for the scholarship previously.

2.5Linear Regression

Linear regression [15]has been used to identify needy students [16]. This model has limitations: it is limited in that it assumes linear relationships between variables under consideration, it is also sensitive to outliers e.g. when a student has characteristics that significantly differfrom other students, the outcome of the evaluation is significantly affected. The model is viewed as being biased in ranking.

Table 1.0: Summary of Manual and Automated Ranking

Model	Context	Strength	Weakness	Limitations
Manual Ranking [10]	Used traditionally where no other option is available	Presents the human touch and ability to interrogate further.	Time consuming, prone to manipulation and bias and human errors	Time consuming, prone to manipulation and errors and objectivity is lost
RFM Model [12]	Grouping based on historical behavior	Can be easily employed to identify needy students	Errors as a result of omission of crucial information	The use of this method tends to have errors as a result of the behaviors of the students.
Multi-Modal Multi Label Approach [13]	Used in determination of the types and amounts of grants	It is easier to acquire the needed information to be used to determine the qualified applicants	The rank operator that is deemed to regularize the prediction results can be biased at certain level.	The rank operator has to regularize the information to ensure modal consistency, and as such certain information may not true in relation to the data collected.
The C4.5 algorithm [14]	Identification of qualified students for the applications	Utilization of information gained ratios	The model relies upon the previous outcome to predict results	Relies on the ratios of information gained to help in predicting the scholarship ranking
linear regression predictor by	Used in identification of needy students	Easy to obtain the predictors by simply measuring	Limited to linear relationships as it assumes	Not appropriate because students differing characteristics,

Gelman and Hill (2007)		household welfare, and using the poverty line to determine the potential benefiters of a social program	straight lines Sensitive to outliers	something that could affect the results.

3.0 METHODOLOGY

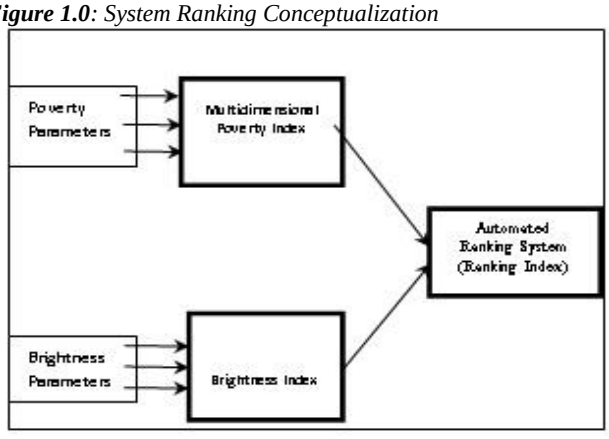
3.1 System Conceptualization and Key steps

The conceptualization of this research was based on reported complains by applicants of scholarship from Constituency Development Fund. Review of available literature on scholarship processing and award, observation and key informant interviews aided in the formulation of the research problem and the definition of the overall objective.The process of the development of the ranking system followed the following steps:

i. The conceptualization of the ranking (**Figure 1.0**)
ii. Review of available literature and open interviews with key informants and stakeholders to identified key parameters that measure level of poverty. (**Table 2.0**)
iii. Review of available literature and open interviews with key informants and stakeholders identified key parameters that measure degree of brightness
iv. Defining the computation of respective indices for "Level of Poverty" and "Degree of Brightness)
v. Defined the computation the overall applicants ranking index.

The System receives input for *Poverty* and *Brightness* parameters and then generates respective indices for Poverty and Brightness and then combines these indices to generate a composite ranking index.

***Figure 1.0**: System Ranking Conceptualization*



3.2 Generation of Ranking Indices

Parameters were identified and utilized to generate indices that were used for ranking of scholarship applicants

a) Parameters for Generating Poverty Level Index[17]

Table 2.0: Poverty Parameters

DIMENSIONS OF POVERTY	INDICATOR	WEIGHT
Health (1/3)	Nutrition	1/6
	Child mortality	1/6
Education (1/3)	Years of schooling	1/6
	School attendance	1/6
Living Standards (1/3)	Cooking fuel	1/18
	Sanitation	1/18
	Drinking water	1/18
	Electricity	1/18
	Housing	1/18
	Assets	1/18
Total		**1.0**

According to the Global Multidimension Poverty Index [17], [18], [19];

$$\boldsymbol{MPI = \text{multidimensional headcount ratio } (H) * \text{intensityofpoverty}(A);}$$

Where;

$$\boldsymbol{H = \left(\frac{q}{n}\right) \ldots\ldots\ldots \text{Eq (i)}}$$

$\boldsymbol{here}$;

$\boldsymbol{q}$ − is the number of people who are multidimensionally poor and

$\boldsymbol{n}$ −is the total population.

$$\boldsymbol{A = \frac{\sum_{i=1}^{n} ci(k)}{q} \ldots\ldots\ldots\ldots \text{Eq (ii)}}$$

Finally, MPI = H*A, (Eq (i) * Eq(ii) and the results is

$$\boldsymbol{Applicant\ Multidemensional\ Poverty\ Index(px) = \frac{\sum_{i=1}^{n} ci(k)}{n} \ldots\ldots\ldots\ldots \text{Eq (iii)}}$$

b) Parameters for Generating Brightness Index

The computation of brightness index was done for two levels;

i. Kenya certificate of Primary Education (KCPE)
ii. Kenya Certificate of Secondary Education(KCSE)

Brightness (Academic Performance) Index for KCPE

The Kenya National Examinations Council (KNEC), therefore, converts all raw marks for each paper at KCPE level using the same mean of 50 and standard deviation of 15 as used in the formula;

Bx= (PI_{KCPE} X_S) = $50 + \left(\frac{XI - \mu}{\delta}\right) + 15$ ……..this equationis standard marks for one KCPE subject while

$\frac{\left(\sum_{i=1}^{n} x_i\right)}{500}$ ……..givestotal summation standard marks for five (5) KCPE subjects taken by a candidate in primary level.

Brightness Performance Index for KCSE

Bx = (PI_{KCSE}) $-\frac{\sum_{i=1}^{n} x_i}{n * 12}$

International Journal of Computer Applications Technology and Research
Volume 13–Issue 05, 35 - 41, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1305.1008

c) Computing the Applicant Ranking Index

Overall Ranking Score (Rx) = MPI*Standardized Students Performance Index

The study adopted this formula to rank applicants where, the overall ranking is the product of Applicant's Multidimensional Poverty Index(MPI) and Applicant's Brightness Performance Index at either KCPE or KCSE.

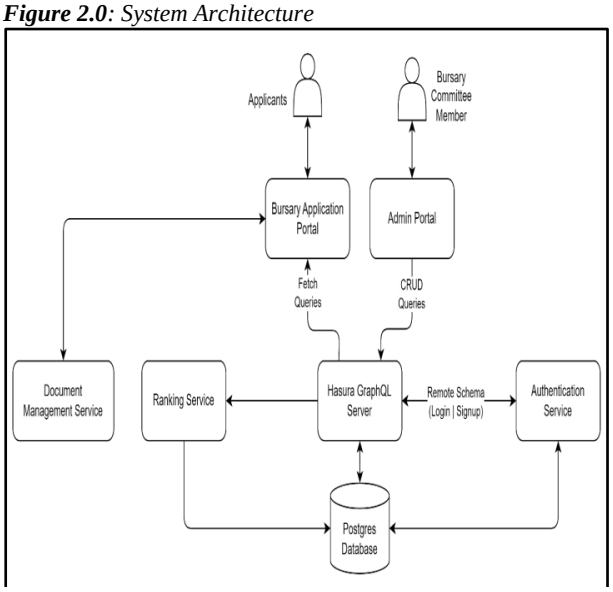
Where:

- $\mathbf{X_s}$ – Standard Marks for the subject
- $\mathbf{X_i}$ – raw marks for the candidate
- **μ** - Mean of raw marks
- **σ** – Standard deviation of raw marks
- **n** – Is number of subjects taken by the students
- ***Bx*** - Applicant's Brightness Performance index in either KCSE orKCPE.
- **PI** - Standardized Students Performance Index
- **12** is the maximum grading system points awarded to grade **A**, - student scores in a subject

3.3 Architecture of the Automated Ranking System

The high level architecture of the system is represented in Figure 2.0

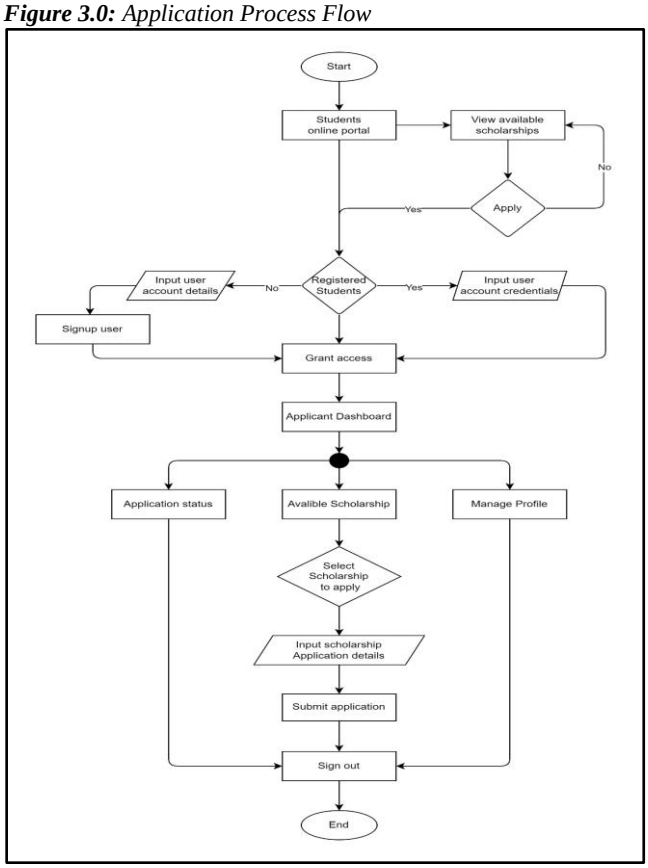
***Figure 2.0**: System Architecture*



3.4 System Logic Flow: Applicant process flow

Figure 3.0 shows the activities performed on the system during the application process and the decision points in the process of ranking.

***Figure 3.0:** Application Process Flow*

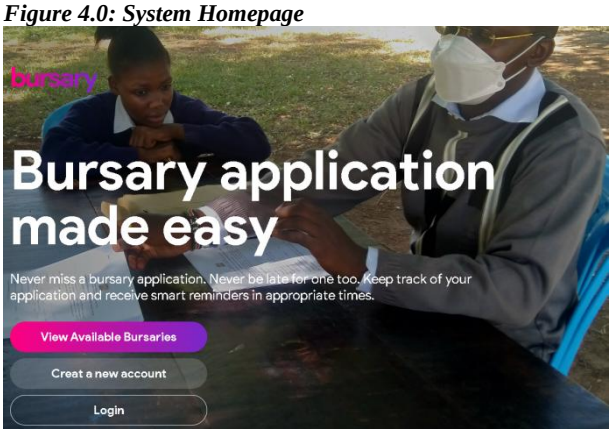


4.0 RESULTS

4.1 System Access and Application Screen

The system was web-based, to allow applicants to make their applications remotely. The homepage allows users to view available scholarships and initiate the application process (Figure 4.0)

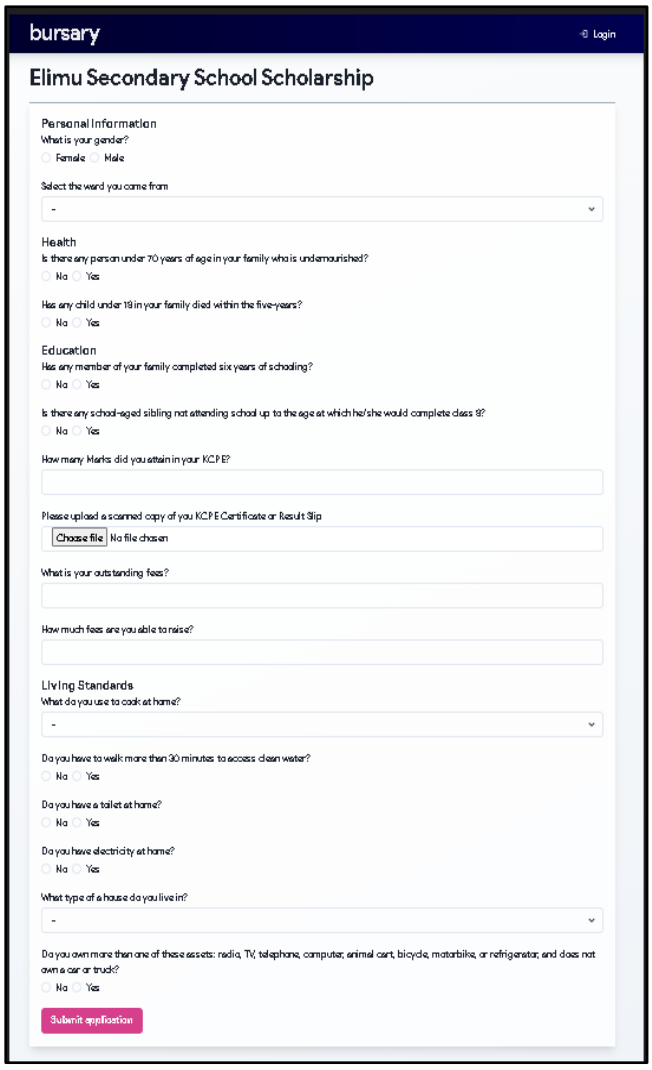
Figure 4.0: System Homepage



Once the user has registered and logged in, the user can start the process of applying for the scholarship by giving details and uploading relevant testimonials (Figure 5.0)

Figure 5.0:Application details capture form

International Journal of Computer Applications Technology and Research
Volume 13–Issue 05, 35 - 41, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1305.1008



4.2Ranking of Applicants

Once the details of all applicants have been received and the testimonials verified, the ranking process is executed. Sample ranking outcome screen is shown in Figure 6.0.

Figure 6.0: System Homepage

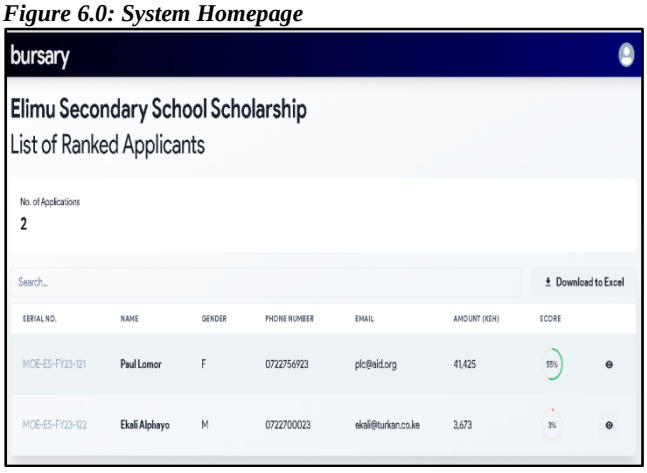


Table 3.0 (*Manual ranking***)** and **Table 4.0 (***System/ automated ranking***)**displays the ranking of the first 20 applicants from a list of 488 applicants. First, it can be seen that the number of applicants that were ranked at the same level differs in the manual and automated ranking. Secondly, the order of ranking of the individuals from the most deserving to the least deserving differs, an indication that the outcomes of manual and system ranking are different.

Table 3.0 *Manual Ranking for KCSE Applicants*

RANKING (ORDER)	APPLICANT ID	GENDER
1	APP169	F
1	APP54	M
1	APP174	F
2	APP7	F
2	APP88	M
2	APP99	M
2	APP101	M
2	APP102	M
3	APP104	M
3	APP167	F
3	APP168	M
3	APP176	M
3	APP78	M
3	APP69	M
4	APP31	F
4	APP10	M
5	APP14	M
5	APP28	M
5	APP51	F
5	APP52	M

Table 4.0 *Manual Ranking for KCSE Applicants*

RANKING (Index)	STUDENT ID	GENDER
0.923	APP1	F
0.923	APP38	M
0.923	APP92	F
0.923	APP93	F
0.923	APP100	F
0.923	APP119	F
0.923	APP19	M
0.923	APP37	M
0.898	APP65	F
0.898	APP101	M
0.898	APP174	F
0.898	APP172	M
0.866	APP2	M
0.866	APP80	F
0.866	APP81	M
0.866	APP82	M
0.866	APP83	M
0.866	APP84	M
0.866	APP85	M
0.866	APP118	M

4.3 Significance of the difference in ranking

The difference between the manually generated ranking and the systems generated ranking was statistically evaluated to determine the level of significance. Chi-square test and the fisher's exact test were utilized. In the test, the null hypothesis ($\mathbf{H_o}$) was "*there exists significant statistical differences between the manual and system generated ranking*", while the alternative hypothesis ($\mathbf{H_1}$) was that "*there exist no significant statistical differences between the manual and system generated ranking*".

Chi-square test computes a P-Value and the outcome are compared to the nominal value of 0.5. When the generated P-Value is less than the nominal value indicates that there is a significant statistical difference between the groups being considered while a value that is greater than the nominal value indicated that there is no significant statistical difference between the groups under consideration.

The results of the Chi-square test yielded a P-value 0.232, a value that is less than the nominal level (0.5) for statistical significance, indicating that significant statistical differences exist between the *manual and system generated rankings* hence the null hypothesis is affirmed.

***Table 5.0**: Chi-Square Test results for Manual (C1) and automated (C2) ranking*

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.110[a]	2	.232
Likelihood Ratio	1.023	2	.261
Linear-by-Linear Association	.037	1	.639
N of Valid Cases	488		

The P-value generated from thefisher exact test was 0.00326, which is less than the nominal value of 0.05. This result indicates that the there is a significant statistical differences between the manual and automated ranking.

***Table 6.0**: Fisher's Test results for Manual (C1) and automated (C2) ranking*

Manual/Automated Ranking	**P-Value**	p-value (two-sided test)	Remark
Manual (C1) & Automated (C2)	0.00163	P=2*0.00163= **0.00326**	**0.00326< 0.05**

5.0 CONCLUSION

The automated ranking of scholarship using standardized parameters generates an outcome that significantly differs from manual ranking, and can thus be considered as a more objective approach.

ACKNOWLEDGMENTS

We are grateful to all participants who played a role in the success of this study. They include; the Constituent Development Committees from the various counties, Ministry of Education officials at local and national levels, Parents to the applicants, successful and unsuccessful applicants who graciously accepted to be part of this study.

REFERENCES

[1] Johnstone, D. B. (2003). Higher Education Finance and Accessibility: Tuition Fees andStudent Lonas in Sub Saharan Africa. Improving Tertiary Education in Sub-SaharanAfrica: Things That Work! Accra.

[2] Gichuhi, L. (2015). Alternative Methods of Financing Higher Education in Kenya . International Journal of Scientific Research and Innovative Technology, 57-67.

[3] Johnstone, B., & Marcucci, P. (2010). Targeting Financial Assistance to Students in Higher Education: Means testing with special emphasis on low and middle income countries. World bank Contract No 0007728373.

[4] Otieno, W. (2009). Higher Education System In Kenya. Kenya.

[5] Centre for International Development. (2010). Constituency Development Funds (CDFs) as a Tool of Decentralized Development. 56th Commonwealth Parliamentary Conference.Nairobi.

[6] Orera, T. M. (2016). Challenges in the disbursement of Constituency Scholarship Fund to Public Secondary School students in Bobasi Constituency Kenya. Nairobi: Kenyatta University.

[7] Ng'alu, M. M., and Bomett, J. E. (2016). The role of constituency development fund in provision of secondary school education in Kenya. International Journal of Scientific and Technology Research 3(8): 429-436.

[8] Siringi, S. (2006). Grand Plan to check high Dropout Rate of poor students. Daily Nation, December 4th, 2006. Nairobi: Kenya

[9] IPAR (2003) . The role of the Higher Education Loans Board in pro-poor management approaches to enhancing access to university education in Kenya

[10] Sifuna,Daniel., (2015), "Development of Education in Africa: The Kenyan Experience". Initiatives Publishers.

[11] Noson S. Y., (2009). "Towards a Definition of an Algorithm". Available on the web at http://arxiv.org/PS cache/math/pdf/0602/0602053v2.pdf. (Last accessed on April 3, 2009.

[12] Dan, Z. (2008). Integrating RFM model and cluster for students loan subsidy Valuation. 2008 International Seminar on Business and Information Management, (pp. 461-464).

[13] Han-Jia Ye, Zhan, D.-C., Li, X., Huang, Z.-C., & Jiang, Y. (2016). College Student Scholarships and Subsidies Granting: A Multi-Modal Multi-Label Approach. IEEE 16th International Conference on Data Mining (pp. 559-568). Nanjing: IEEE.

[14] Ruggieri S, (2002). Efficient C4.5 (classification algorithm), IEEE Transactions on Knowledge and Data Engineering, 14(2), pp. 438–444.

[15] Gelman and Hill (2007)Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press

[16] Castano, E. (2002). Proxy Means Test Index for targeting social programs: Two Methodologies and Empirical Evidence. Lecturas de economia, no 56.

[17] Alkire, S., Kanagaratnam, U. and Suppa, N. (2020). 'The global Multidimensional Poverty Index (MPI): 2020

revision', OPHI MPI Methodological Note 49, Oxford Poverty and Human Development Initiative, University of Oxford.

[18] Kenya National bureau of statistics (2016), Ministry of Planning and National Development. basic Report on Well-being in Kenya, based on Kenya Integrated household budget survey- 2015/16

[19] Mwai, N.,and Hasan, B., (2007). The Kenya National Examinations Council KCPE Examination report.