Recommender System For Stem Enrolment In Universities Using Machine Learning Algorithms: Case Of Kenyan Universities

Benard Ondiek[1]*, Lucy Waruguru[2], Stephen Njenga[3]

[1,2] School of Technology, KCA University, Nairobi, Kenya
[3] Department of Computer Science, Murang'a University of Technology, Muranga, Kenya
*Corresponding Author:
Email: benmacondiek@gmail.com

***Abstract**.*

Technology, Engineering, and Mathematics (STEM) enrolment has gained a lot of research interest. The increase in demand for STEM-based skill sets has contributed to the need for systems that could potentially increase enrolments in the field. The purpose of this study was to investigate recommender systems for STEM enrolment in universities using machine learning algorithms. Students face challenges while selecting STEM courses that match their attributes. This article aims to provide a recommender system for STEM enrolment using machine learning algorithms. The article investigates three machine learning algorithms which include Support Vector Machine (SVM), Artificial Neural Network (ANN), and Naïve Bayes. Accuracy and validation techniques were applied to test the algorithms. The results demonstrated that our work performed better than that of the published research, with the ANN outperforming other classification methods. The results position ANN as an important algorithm in building a recommender model for STEM higher education enrolment. The study also identifies high school grades and Interest in STEM courses as important features in predicting STEM course enrolment in higher education. The study will guide policy on the courses to lay more emphasis on, as well as for the funding authorities to prioritize funding allocation for STEM-based courses.

Keywords: *Machine Learning, Higher Education, Support Vector Machine, Engineering and Mathematics.*

I. INTRODUCTION

Despite the need for a more skilled workforce, STEM has not attracted many students. This has been attributed to the challenge of misalignments of students' capabilities with the courses they choose in Higher Education [1]. According to He and Jang [2], there is a requirement for national efforts to establish STEM education due to the worldwide desire for STEM workers. Technological innovation and the search for "need-based and practical solutions" drive the four distinct STEM disciplines" [3]. As presented by Sustainable Development Goal(SDG) 4 [4] there is a drive for students to select courses that will result in more innovations. As stated by [5], researchers are actively looking for ways to improve STEM courses since they believe these courses can help students develop "21st-century skills"[6]. This has been reinforced by Tawbush et al. [7] who stress that researchers are increasingly concentrating on STEM because of the development of technology[8] and the rise in environmental threats. The growth experienced in STEM courses would be attributed to the need to reduce the gap between academia and the industry [9]. Science, Technology, Engineering and Mathematics comes in handy to propel the economy in the ever-increasing competition [3], [10]. Despite the numerous choices a student must make while selecting courses, getting the appropriate courses is still daunting [11]. A recommender system can be defined as a platform that helps match students' interests with the STEM courses they are applying to [12], [13]. According to Girase et al.[14], the Recommender system encompasses several artificial intelligence techniques. Some techniques widely used with recommender systems include machine learning and data mining [15].

Recommendation comprises three phases which include information, the learning phase, and finally the prediction [16]. Recommender systems' primary goal is to reduce users' information overload while allowing the personalization of recommendations for better decision-making[17]. Recommender Systems are useful for giving students advice while choosing disciplines to study in higher education[18]. The outcome is improved academic performance motivated by a reduction in dropout rates [19]. Recommendation systems can be classified into content-based, collaborative, and hybrid recommender systems [20]–[22]. Advances have increased, especially in the development of recommender systems for enrolment in higher education. Markov chain model [23], [24], Regression[25], and analytical hierarchy process [26], [27] have been used in the development of recommender systems in higher education.In the recent past research has drifted to more advanced approaches in recommending course enrolments in Higher education. The approaches include Datamining [28], Machine learning[29], and Deep Learning [30], [31]. The goal of machine learning, a subfield of artificial intelligence, is to produce algorithms based on "data trends and historical relationship between data"[32], [33]. Supervised learning[34] and unsupervised learning [35] are the two main classifications of machine learning algorithms [36]–[38]. According to [37] Supervised learning utilizes “labeled datasets” which conduct training to the algorithms to classify data or predict outcomes accurately. On the other hand, unsupervised algorithms can “discover hidden patterns” without getting help from humans [39], [40].

The study will build recommendations for STEM courses in higher education based on survey data from students who are currently in Kenyan Universities. To achieve our objective, we train three machine learning models (SVM, Naïve Bayes, and ANN) based on the student’s historical data.Several communities and organizations are seeking ways in which they can provide education for all in line with SDG4. The main focus for SDG 4 is innovative learning which STEM can articulate. Sithole et al. [41] argue that Universities are facing the challenge of low STEM enrolments while at the same time having high attrition rates in STEM enrolment. According to the Commission of higher education [42] in their report for 2017/2018 statistics, University enrolments are weighing more on the “arts, humanities and social sciences” compared to the STEM courses. The report further states that the STEM-related courses enrolment was at 41% compared to the art, humanities, and social sciences which were rated at 59%.The research aims at implementing a recommender system that will attempt to increase STEM enrolments. This will be achieved by enabling the students to get enrolled in a course based on their interests and motivations, socio-demographics, academic strengths, and educational factors. There have been attempts to develop recommender models by researchers in the past.

Mokarrama et al. [43] proposed a content-based recommender system for selecting universities. Based on a set of features, the system was able to suggest Universities for the students to select. The challenge with the system was the scope [44] which was more general and the prospective student’s preferences were not well captured. They did not address the Kenyan domain in the study. Wang et al. [39] developed a model based on machine learning and hybrid techniques. The model was able to recommend courses for students in successive semesters. The model implemented the matrix factorization machine learning algorithm. The system had a recall percentage of 58% and an accuracy of 78%. The weakness of the model was the challenge of information on demographics and past grades as suggested by Wang [39]. The study will seek to include demographic details and previous grades as suggested by [39]. The study will also include student preference while being specific to Kenyan Universities and STEM domain.This article aims at investigating the three-machine learning algorithms that are used in recommending STEM courses to students joining University. The article also aims at identifying the machine learning algorithm with the highest level of accuracy and the features that are best suited to create a recommendation. The current study investigated the following research questions.Question 1: Can we model the STEM recommender path choice according to the student’s academic history by applying different ML algorithms?
Question 2: Which ML classifier offers optimal performance in predicting student STEM course selection?
Question 3: How is a student’s STEM path choice associated with that student’s previous academic performance?

This article is organized as follows: Section 2 discusses related work; section 3 explains the materials and methods, and section 4 discussion of the results. Section 5 is the last section that contains the conclusions and future works.

II. LITERATURE REVIEW

Several studies have been conducted in the area of course recommendation using machine learning algorithms in higher education. The most popular Machine learning algorithms used in higher education recommendation include SVM, Naïve Bayes, and Artificial Neural networks.

State of the art in Machine Learning Recommender Model in higher education

Pupara et al.[45] developed an institution-based recommender system that was based on the student's characteristics. The system relied on the student's attitude and characteristics to recommend institutions of learning. The study used 1109 records of students from three Universities within Pakistan. The study implemented the decision tree model and association mining rule. The accuracy of the model was 69% based on features that included: the university's reputation, public confidence in institutions, student skills, and family income [44]. The research was limited to recommendation courses based on attitudes and applicants' characteristics. Baskota and Ng[11], developed a machine learning recommender system for graduate school recommendation using the SVM and K Nearest Neighbors. The model used personal data while at the same time utilizing data obtained from online portals. The SVM model was implemented to find an appropriate graduate school whereas the KNN was implemented to get any other university that the student would be enrolled in based on their profiles. The model achieved an accuracy of 61%. The model agreed with the graduate recommender model that was proposed by Aishwarya and Tiple[16]. The limitation of the study was its application to the graduate student's enrolment. Fiarni et al.[46] developed a machine learning recommender system that predicted the information system majors based on student's interest and performance.

The system implemented the decision tree machine learning algorithm. The system performed well by posting a precision of 71% and a Recall of 61%. The system was not evaluated using the accuracy technique thereby not showing the robustness of the model compared to other similar ones.Alsayed et al.[47] developed a system that recommended majors for students in higher education. The study implemented several supervised models that included decision trees, random forests, gradient boosting classifiers, and extra tree classifiers. The dataset was obtained from Kaggle online database. The model was validated using the 10-fold cross-validation method which achieved an accuracy of 75% and 61% for Random forest and gradient-boosting classifiers. According to nam the shortcoming of the model included the use of the Kaggle dataset which did not have customized input features. The study further proposed the use of survey data to enhance the accuracy of the prediction.Jena et al.[48] developed eLearning courses recommender model that uses SVD and KNN. The model aimed at recommending eLearning courses to students while achieving the highest accuracy. The Accuracy of the KNN outperformed the other models. The model was built on the Python environment.

The limitation of the study was the application to online students rather than the whole student fraternity.Zayed et al.[49] improved the intelligent recommender system developed by Alsayed et a.l [47]. The study implemented supervised machine learning algorithms that include Support Vector Machine, Random Forest, and Decision Trees. The research used the same Kaggle dataset but included hyper-tuning to increase the accuracy. According to Zayed et al.[49] the accuracy of the random forest increased to 95%. The main limitation of the study was the inability to get the input features that would fit the problem.Despite the concentration of research on machine learning recommender models in higher education, there is little evidence that the researchers have concentrated on STEM education using primary data. To the best of our knowledge, no research has been conducted on machine learning recommender systems for STEM courses in Kenya. This gives provides the need for research which will have implications for higher education, students, and other education stakeholders.

Machine learning models for recommending enrolment in higher education.

Support Vector Machine

SVMs' popularity has been on the rise due to their robustness [50]. Srivastava and Bhambhu, [51] define SVM as a set of related supervised learning methods implemented in both classification and regression According to Baskota and Ng[11], [52], [53] SVM has been more instrumental in text classification due to their ability to "classify both linear and non-linear data"[54]. Despite their low speed as far as training time is concerned they attain a high level of accuracy compared to other algorithms [55]. According to Srivastava and Bhambhu [51] SVM has a high accuracy level compared to other classification algorithms. The SVM algorithm is credited for having been used in various fields which include "text categorization, image classification [56] and object detection and data classification" [51].

According to Roy and Dutta[13] SVM would be appropriate in solving the challenge of high dimensional problems that arise in recommender systems. The performance of SVM is stable with or without the addition of new data. Below is a representation of the model.

Artificial Neural Network

ANN is a type of Artificial intelligence technique whose main strength is the ability to self-learn and generate efficient results [57], [58] It is independent of data types thereby being able to learn patterns independently [59].According to Hernandez et al.[60] an ANN system processes information based on units that are referred to as neurons. ANN draws its strength from the idea of the human brain works. The brain is interconnected to various neurons to make decisions faster [59]. According to Hernandez et al. [60] the strength of the ANNs is the ability to perform well in terms of metrics. This is in comparison with the other classification algorithms [61]. According to Latifah et al.[62], Artificial neural networks' plasticity, nonlinearity, modularity, and openness to noisy, fuzzy, or soft data give it an edge over other algorithms.

Naïve Bayes

Naïve Bayes is widely acceptable due to its simplicity and speed of performing tasks [63]. Naive Bayes is based on the Bayesian principle of the theory of probability[34]. The operation of the Naïve Bayes is based on a “subjective probability of some unknown states under incomplete information” [64].

III. MATERIALS AND METHODS

The study contributes to the body of knowledge by building a machine learning recommender model for STEM enrolment in Kenya Universities. The process of development of the machine learning recommender model comprises three stages. The stages include data collection, data pre-processing, and visualization and recommendation as shown in Figure 1 below.

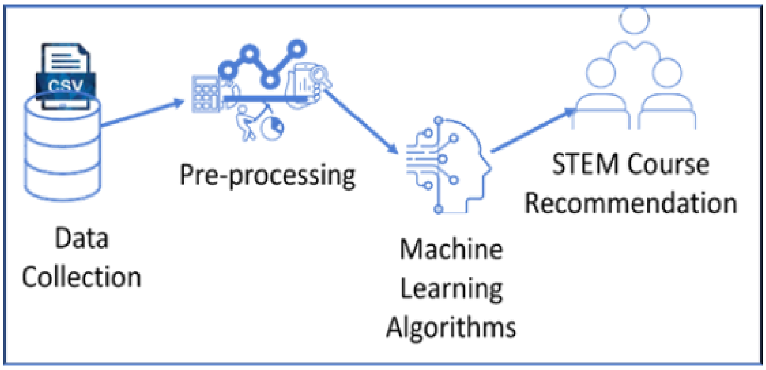

Fig 1. Workflow of machine learning algorithms[Author:The researcher]

Data Collection and Pre-processing

The data used in the study was collected through a survey conducted among students within four Universities in Kenya. The universities were categorized into both private and public. The targeted Universities had students who were taking STEM-based courses and had previously attended the Kenya Certificate of Secondary Education Exam (KCSE). The data contained 384 sample records with 38 input features and was based on the stratified random sampling technique. The technique brings the strength of selecting the sample properly enabling the researcher to generalize the findings for the entire population [35].

The dataset comprises features that include but are not limited to gender, age bracket, family income, KCSE final score, and score for individual STEM-based courses. Our dataset comprises both numerical and categorical data. The feature selection was conducted using the lasso technique [65]) which resulted in 29 features. By combining the advantages of ridge regression with subset selection, LASSO enhances model interpretability and prediction accuracy [65]. The records were then augmented using the smote technique to 1158. The STEM courses were labeled as 0, 1, 2, and 3 representing Engineering, Mathematics, Sciences, and Technology respectively.The research was based on survey data due to the need of using primary data. The choice of the dataset was a shift from the traditional choice of data that resides within the local university repositories as suggested by Alzayed et al.[47].

Data Processing and Visualization

In this stage, the data undergoes a cleaning process to have it ready for visualization and training process. The cleaning was conducted using the Python library [48]. The label encoder was then implemented to transform the categorical features into numeric values.

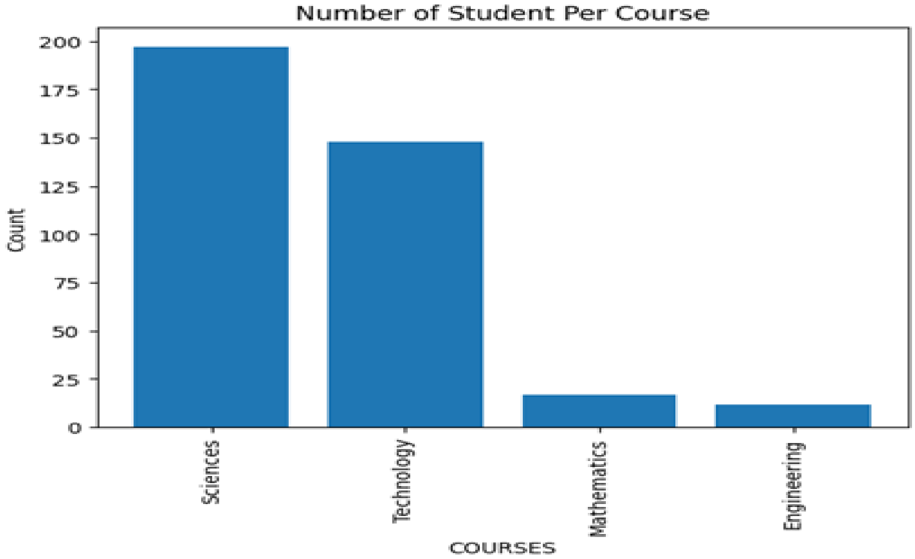

Fig 2. Distribution of STEM courses within selected universities

Source: Field Data 2023

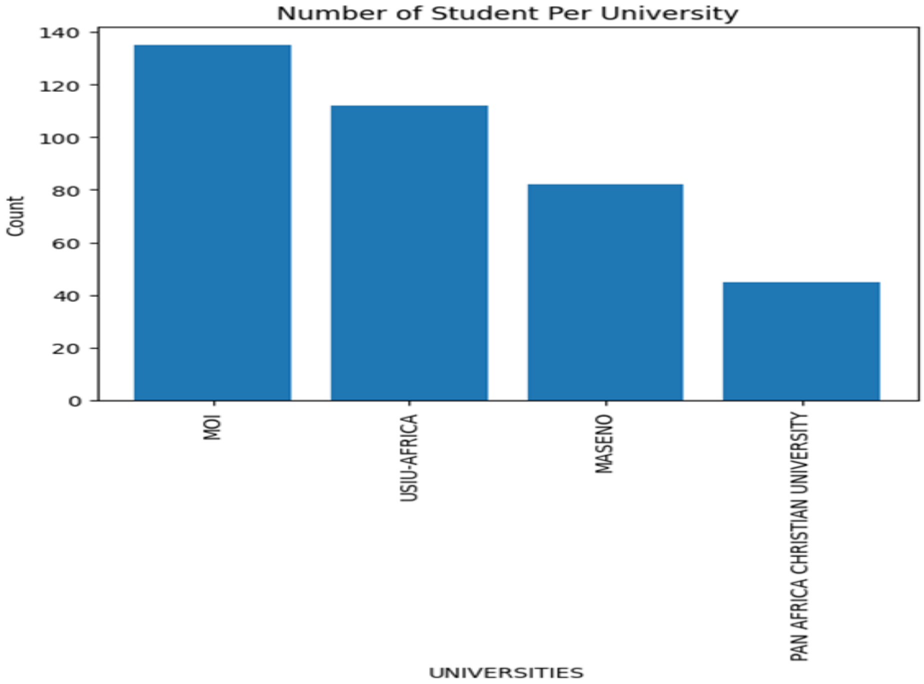

Fig 3. Distribution of students per University

Source: Field Data 2023

Model selection

The training and testing of the model are based on three Machine learning algorithms namely SVM, Naïve Bayes, and Artificial Neural Network with the help of the 10-fold cross-validation technique. This is achieved through the use of data that has undergone pre-processing using the Python library known as Scikit-learn [48] library. The data is split into training and testing data using the ratio 70:30.After training, the model will either determine the best model on the current study dataset or identify patterns between input characteristics and output variables. Accuracy will be used to identify the high-performance machine learning algorithm. Finally, ML models with high accuracy are selected for model development.

IV. RESULT AND DISCUSSION

The selection of the appropriate STEM courses is an important component for higher education stakeholders. This section investigates STEM recommendations using machine learning algorithms in Kenyan universities. Python libraries as well as visualizations are implemented to answer the research questions.

Question 1: Can we model the STEM recommender path choice according to the student's academic history by applying different ML algorithms?

Question 2: Which ML classifier offers optimal performance in predicting student STEM course selection?

The first two questions were comprehensively addressed by conducting experiments for the three machine learning models.

4.1 SVM

SVM is a popular machine-learning algorithm that works best with small datasets [11]. The model is trained and tested as shown in Figure 5 below.

The accuracy of the SVM is 77.22%. This is in line with the previous study by Zayed et al.[49]. The high performance is also supported by Gaye et al.[55], [66] who assert that despite the high duration required to train the SVM model, they post very good performance. The Confusion matrix was also used to test the number of positively predicted STEM courses as shown in figure 6 below.

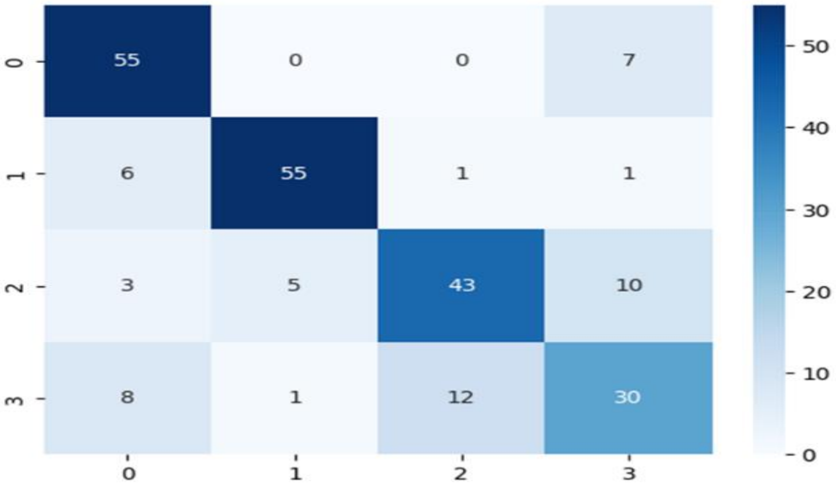

Fig 4. SVM confusion matrix.
Source: Field Data 2023

According to Kulkarni et al.[67], a confusion matrix is a predictive tool in machine learning that is deployed to test the count of predicted and actuals. The results suggest that 55 engineering students were recommended correctly based on the total number of 72 students. This represents a percentage of 76 accurate recommendations for engineering students. Its also worth noting that 55 mathematics students were also predicted accurately from the 61 students. This represents 90% of positive recommendations from the test data. There were 43 accurate predictions for the sciences out of the possible 56. This represents 77% accurate predictions. Finally, technology was presented with 30 accurate predictions which was a 44% accurate prediction. The low number of predictions on the technology courses is due to the level of education that most technological students are taking. From the data, most students taking technology were certificate students who didn't have the requisite qualification to be placed in a University.

4.2 Naïve Bayes

Simplicity and speed of task completion are the driving factors for the Naïve Bayes machine learning algorithm [63]. The accuracy of the Naïve Bayes is 72.22%. The model is trained and tested as shown in Figure 5 below.

The below image shows the confusion matrix after training the STEM recommendation based on the Naïve Bayes algorithm.

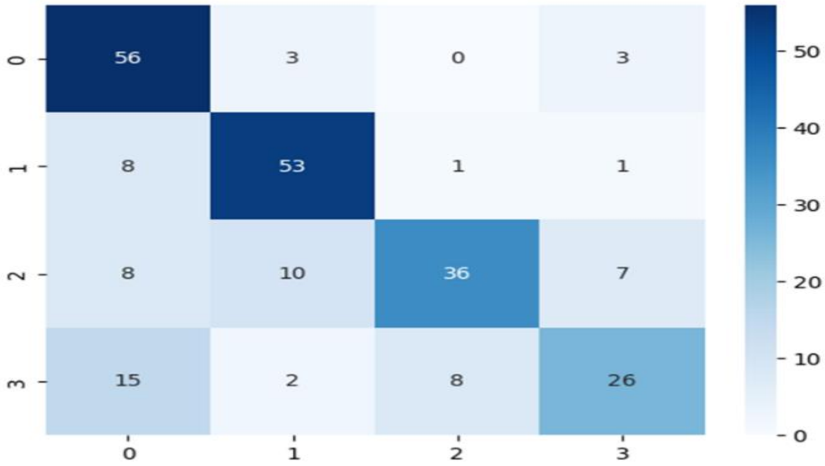

Fig 5. Naïve Bayes model confusion matrix
Source: Field Data 2023

According to the confusion matrix, 64% of the predictions for the Engineering courses were accurate. Mathematics posted 78% accuracy in predictions while science and technology posted 80% and 70% respectively.

4.3 ANN

Latifah et al. [62] argue that ANN performs well compared to other algorithms. ANN draws its strength from the idea of the human brain works. The brain is interconnected to various neurons to make decisions faster [59]. The study presents training of the model using the ANN machine learning model as shown in Figure 9. The accuracy posted by ANN is 82%. This is the best result in comparison to both Naïve Bayes and SVM.Figure 6 shows the confusion matrix after training the STEM recommendation based on the ANN algorithm.

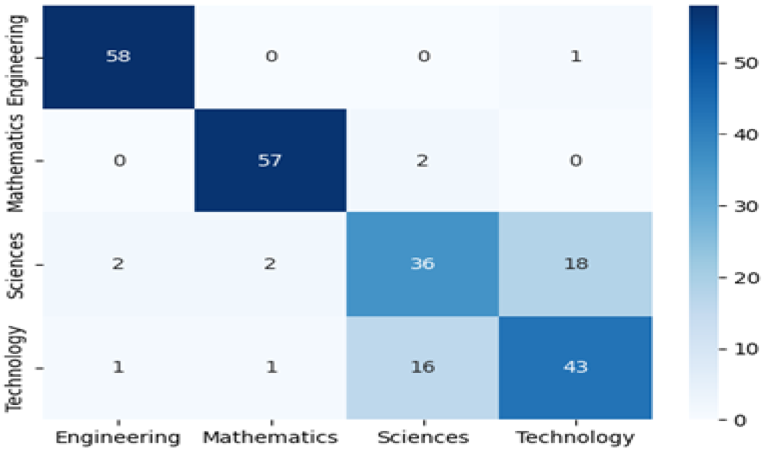

Fig 6. ANN Model confusion matrix
Source: Field Data 2023

According to the confusion matrix, 92% of the predictions for the Engineering courses were accurate. Mathematics posted 95% accuracy in predictions while science and technology posted 66% and 70% respectively. The observations agree with [13] whose findings explain that ANN produces high performance. Question 3: How is a student's STEM path choice associated with that student's previous academic performance? To answer the second research, question the research used the Spearman correlation model which was conducted within the Python environment. Since the Spearman correlation is discovered by computing the Pearson correlation on the ranked data inside two features, it is often referred to as Spearman's "rank correlation" coefficient [68]. The Spearman correlation presented the relationship between the features. The strongest positive and negative correlations are 1 and 1, respectively, in the coefficient's range of 1 to 1 [69]. Figure 11 below shows how the features are related.

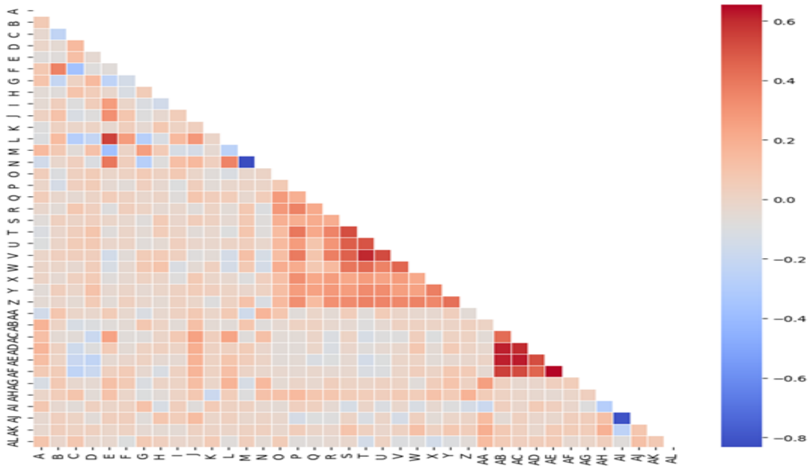

Fig 7. Relationship of the features
Source: Field Data 2023

According to the model, there is a strong positive correlation between feature J and features AB, AC, AD, and AE. These are the features that represent STEM courses about the courses that the participants are currently enrolled in. This historical information is important for recommending courses to future students [13]. The results suggest that students who performed well in the sciences i.e. Biology Physics and Chemistry are more likely to join sciences in universities in Kenya. Similarly, students who posted high scores in Mathematics are likely to join courses in Statistics and other mathematics-related courses. The students who posted good scores in Mathematics, Physics, and Chemistry are more likely to join the Engineering courses in Kenyan Universities.

The study was also able to identify that it was difficult to predict the technological students due to the progression from certificate to bachelor. Most of the sampled students taking the technological courses had performed poorly in the STEM courses in secondary but managed to progress to undergraduate due to course progression. This was confirmed by Pearson’s correlation and the various confusion matrix presented. The study also showed that peer influence and interest in STEM courses affect the student’s choice of STEM courses. This is evident from the strong positive correlation between the Peer influence feature and the course taken. This agrees with [70]) whose research indicated that a student who has an interest either by own violation or the impact of peers would more likely select a STEM course in university. This is also supported by [61] who assert that awareness of career would most like influence a student’s choice of STEM course.

V. CONCLUSION

To achieve higher education accessibility in the agenda as per SDG 4, it’s important to match the student’s interest with the courses offered. This ensures reduced dropout rates in Universities while increasing retention and graduation rates. Machine learning recommender models have attracted a lot of interest in research due to their predictive and recommendation prowess. Limited studies have been conducted in the area of recommendation within the STEM higher education domain using survey data. The study implemented the ANN, Naïve Bayes, and SVM algorithms.The study sampled 4 universities in Kenyan Universities that offered STEM courses based on the stratified random sampling technique. The Universities were classified into both public and private. The data was collected through questionnaires that were administered through google forms and controls were put in place to prevent missing data.The analysis was done on the Python environments and results were presented using data visualization libraries.

The accuracy of the three algorithms was tested with ANN giving the highest performance. Confusion Matrix was generated on the three algorithms to present the level of accurate recommendations for STEM courses. Finally, the relationship between the features was presented and the results suggested a strong positive correlation between the scores for STEM courses in High school and the Enrolled course. Similarly, there is a strong correlation between the interest in STEM courses and the courses that are enrolled.Based on the Lasso for feature selection. method previous KCSE score and the exposure to STEM-related activities are good criteria for suggesting student field specialization. In addition, these ML models and features could be of high value in developing a system to easily recommend a STEM course to potential applicants who are often uncertain of their desired fields of specialization. Finally, this article differs from other research in this area of predicting student courses in a university because it is based on the Kenyan context and the area specific to STEM enrolment

VI. RECOMMENDATION

The study serves as a basis for the investigation of machine learning models in recommender systems in higher education. Future studies should be conducted on the application of deep learning algorithms in the recommendation of STEM courses within the higher education domain.

REFERENCES

[1] S. Kaleva, J. Pursiainen, M. Hakola, J. Rusanen, and H. Muukkonen, “Students’ reasons for STEM choices and the relationship of mathematics choice to university admission,” ***Int. J. STEM Educ.***, vol. 6, no. 1, p. 43, Dec. 2019, doi: 10.1186/s40594-019-0196-x.

[2] Z. He and W. Jiang, “A new belief Markov chain model and its application in inventory prediction,” ***Int. J. Prod. Res***., vol. 56, no. 8, pp. 2800–2817, Apr. 2018, doi: 10.1080/00207543.2017.1405166.

[3] J. Sharma and P. K. Yarlagadda, “Perspectives of ‘STEM education and policies’ for the development of a skilled workforce in Australia and India,” ***Int. J. Sci. Educ.***, vol. 40, no. 16, pp. 1999–2022, Nov. 2018, doi: 10.1080/09500693.2018.1517239.

[4] C. Kroll, A. Warchold, and P. Pradhan, “Sustainable Development Goals (SDGs): Are we successful in turning trade-offs into synergies?,” Palgrave Commun., vol. 5, no. 1, p. 140, Dec. 2019.

[5] F. Liu, “Addressing STEM in the context of teacher education,” J. Res. Innov. Teach. Learn., vol. 13, no. 1, pp. 129–134, Jan. 2020, doi: 10.1108/JRIT-02-2020-0007.

[6] S. Fajrina, L. Lufri, and Y. Ahda, “Science, Technology, Engineering, and Mathematics (STEM) as A Learning Approach to Improve 21st Century Skills: A Review,” Int. J. Online Biomed. Eng. ***IJOE***, vol. 16, no. 07, p. 95, Jun. 2020, doi: 10.3991/ijoe.v16i07.14101.

[7] R. L. Tawbush, M. A. Webb, S. D. Stanley, and T. G. Campbell, “International comparison of K-12 STEM teaching practices,” J. Res. Innov. Teach. Learn., vol. 13, no. 1, pp. 115–128, Jan. 2020, doi: 10.1108/JRIT-01-2020-0004.

[8] M. F. Kamaruzaman, R. Hamid, A. A. Mutalib, and M. S. Rasul, “Comparison of Engineering Skills with IR 4.0 Skills,” Int. J. Online Biomed. Eng. IJOE, vol. 15, no. 10, p. 15, Jun. 2019, doi: 10.3991/ijoe.v15i10.10879.

[9] B. Freeman, S. Marginson, and R. Tytler, “An international view of STEM education,” 2019. doi: 10.1163/9789004405400_019.

[10] D. Ardianto, H. Firman, A. Permanasari, and T. Ramalis, What is Science, Technology, Engineering, Mathematics (STEM) Literacy? 2019. doi: 10.2991/aes-18.2019.86.

[11] A. Baskota and Y.-K. Ng, “A Graduate School Recommendation System Using the Multi-Class Support Vector Machine and KNN Approaches,” in 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT: IEEE, Jul. 2018, pp. 277–284. doi: 10.1109/IRI.2018.00050.

[12] H. A. Yazdi, S. J. S. M. Chabok, and M. Kheirabadi, “Dynamic Educational Recommender System Based on Improved Recurrent Neural Networks Using Attention Technique,” Appl. Artif. Intell., pp. 1–24, Dec. 2021.

[13] D. Roy and M. Dutta, “A systematic review and research perspective on recommender systems,” ***J. Big Data***, vol. 9, no. 1, p. 59, May 2022, doi: 10.1186/s40537-022-00592-5.

[14] S. Girase, V. Powar, and D. Mukhopadhyay, “A user-friendly college recommending system using user-profiling and matrix factorization technique,” in 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida: IEEE, May 2017, pp. 1–5. doi: 10.1109/CCAA.2017.8229779.

[15] K. J. Singh, D. S. Kapoor, and B. S. Sohi, “All about human-robot interaction,” in Cognitive Computing for Human-Robot Interaction, Elsevier, 2021, pp. 199–229. doi: 10.1016/B978-0-323-85769-7.00010-0.

[16] N. Aishwarya and B. Tiple, “The University Recommendation System for Higher Education,” ***Int. J. Recent Technol.*** Eng., vol. 8, no. 6, pp. 1692–1696, Mar. 2020, doi: 10.35940/ijrte.F7632.038620.

[17] D. Ferreira, S. Silva, A. Abelha, and J. Machado, “Recommendation System Using Autoencoders,” Appl. Sci., vol. 10, no. 16, 2020, doi: 10.3390/app10165510.

[18] Sciforce, "medium.com," Deep Learning Based Recommender Systems. Accessed: Mar. 14, 2022.

[19] T. N. D. Oliveira, F. Bernardini, and J. Viterbo, "An Overview on the Use of Educational Data Mining for Constructing Recommendation Systems to Mitigate Retention in Higher Education," in 2021 IEEE Frontiers in Education Conference (FIE), Oct. 2021, pp. 1–7. doi: 10.1109/FIE49875.2021.9637207.

[20] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, "Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis," Front. Psychol., vol. 12, p. 698490, Dec. 2021, doi: 10.3389/fpsyg.2021.698490.

[21] Z. Gulzar, A. Leema, and G. Deepak, "PCRS: Personalized Course Recommender System Based on Hybrid Approach," Procedia Comput. Sci., vol. 125, pp. 518–524, Jan. 2018, doi: 10.1016/j.procs.2017.12.067.

[22] L. Guo, J. Liang, Y. Zhu, Y. Luo, L. Sun, and X. Zheng, "Collaborative filtering recommendation based on trust and emotion," ***J. Intell. Inf. Syst***., vol. 53, no. 1, pp. 113–135, Aug. 2019, doi: 10.1007/s10844-018-0517-4.

[23] V. O. Ezugwu and S. Ologun, "Markov chain: a predictive model for manpower planning," ***J. Appl. Sci. Environ. Manag.***, vol. 21, no. 3, p. 557, Jul. 2017, doi: 10.4314/jasem.v21i3.17.

[24] A. Polyzou, A. N. Nikolakopoulos, and G. Karypis, Scholars Walk: A Markov Chain Framework for Course Recommendation. 2019.

[25] K. Kumari and S. Yadav, "Linear regression analysis study," ***J. Pract. Cardiovasc. Sci***., vol. 4, p. 33, Jan. 2018, doi: 10.4103/jpcs.jpcs_8_18.

[26] K. B. Sangka and B. Muchsini, "Accommodating Analytic Hierarchy Process (AHP) for Elective Courses Selection," IJIE Indones. ***J. Inform. Educ.***, vol. 2, no. 2, Dec. 2018, doi: 10.20961/ijie.v2i2.24436.

[27] H. Liang, J. Ren, S. Gao, L. Dong, and Z. Gao, "Chapter 8 - Comparison of Different Multicriteria Decision-Making Methodologies for Sustainability Decision Making," in Hydrogen Economy, A. Scipioni, A. Manzardo, and J. Ren, Eds., Academic Press, 2017, pp. 189–224. doi: 10.1016/B978-0-12-811132-1.00008-0.

[28] M. Ye, "The Datamining Algorithm on Knowledge Dependence," in 2018 International Conference on Smart Grid and Electrical Automation (ICSGEA), Changsha: IEEE, Jun. 2018, pp. 234–236. doi: 10.1109/ICSGEA.2018.00065.

[29] I. E. Guabassi, Z. Bousalem, R. Marah, and A. Qazdar, "A Recommender System for Predicting Students' Admission to a Graduate Program using Machine Learning Algorithms," ***Int. J. Online Biomed***. Eng. IJOE, vol. 17, no. 02, p. 135, Feb. 2021, doi: 10.3991/ijoe.v17i02.20049.

[30] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," ***J. Big Data,*** vol. 2, no. 1, p. 1, Dec. 2015, doi: 10.1186/s40537-014-0007-7.

[31] R. Vargas, A. Mosavi, and R. Ruiz, "DEEP LEARNING: A REVIEW,"Adv.Intell.Syst.Comput.,vol.5, Jun. 2017.

[32] S. Angra and S. Ahuja, "Machine learning and its applications: A review," in 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Mar. 2017, pp. 57–60.

[33] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," Electron. Mark., vol. 31, no. 3, pp. 685–695, Sep. 2021, doi: 10.1007/s12525-021-00475-2.

[34] I. E. Guabassi, Z. Bousalem, R. Marah, and A. Qazdar, "Comparative Analysis of Supervised Machine Learning Algorithms to Build a Predictive Model for Evaluating Students' Performance," ***Int. J. Online Biomed. Eng. IJOE,*** vol. 17, no. 02, p. 90, Feb. 2021, doi: 10.3991/ijoe.v17i02.20025.

[35] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Comput. Sci., vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.

[36] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, "Introduction to Machine Learning, Neural Networks, and Deep Learning," Transl. Vis. Sci. Technol., vol. 9, no. 2, pp. 14–14, Feb. 2020, doi: 10.1167/tvst.9.2.14.

[37] P. Louridas and C. Ebert, "Machine Learning," IEEE Softw., vol. 33, no. 5, pp. 110–115, Sep. 2016, doi: 10.1109/MS.2016.114.

[38] P. Ayush, "towardsdatascience.com," Introduction to Machine Learning for Beginners. [Online]. Available: https://towardsdatascience.com/introduction-to-machine-learning-for-beginners-eed6024fdb08

[39] X. Wang, "Course-Taking Patterns of Community College Students Beginning in STEM: Using Data Mining Techniques to Reveal Viable STEM Transfer Pathways," Res. High. Educ., vol. 57, no. 5, pp. 544–569, Aug. 2016, doi: 10.1007/s11162-015-9397-4.

[40] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, "A survey on missing data in machine learning," J. Big Data, vol. 8, no. 1, p. 140, Oct. 2021, doi: 10.1186/s40537-021-00516-9.

[41] A. Sithole, E. T. Chiyaka, P. McCarthy, D. M. Mupinga, B. K. Bucklein, and J. Kibirige, “Student Attraction, Persistence and Retention in STEM Programs: Successes and Continuing Challenges,” High. Educ. Stud., vol. 7, no. 1, p. 46, Jan. 2017, doi: 10.5539/hes.v7n1p46.

[42] “CUE,” CUE. Accessed: Jul. 16, 2022. [Online]. Available: https://www.cue.or.ke/

[43] M. Mokarrama, S. Khatun, and M. Arefin, “A content-based recommender system for choosing universities,” Turk. J. Electr. Eng. Comput. Sci., vol. 28, pp. 2128–2142, Jul. 2020, doi: 10.3906/elk-1911-37.

[44] N. C. Siregar and R. Rosli, “The effect of STEM interest base on family background for secondary student,” ***J. Phys. Conf. Ser.,*** vol. 1806, no. 1, p. 012217, Mar. 2021, doi: 10.1088/1742-6596/1806/1/012217.

[45] K. Pupara, W. Nuankaew, and P. Nuankaew, “An institution recommender system based on student context and educational institution in a mobile environment,” in 2016 International Computer Science and Engineering Conference (ICSEC), Dec. 2016, pp. 1–6. doi: 10.1109/ICSEC.2016.7859877.

[46] C. Fiarni, E. M. Sipayung, and P. B. T. Tumundo, “Academic Decision Support System for Choosing Information Systems Sub Majors Programs using Decision Tree Algorithm,” ***J. Inf. Syst. Eng. Bus. Intell***., vol. 5, no. 1, p. 57, Apr. 2019, doi: 10.20473/jisebi.5.1.57-66.

[47] A. O. Alsayed et al., “Selection of the Right Undergraduate Major by Students Using Supervised Learning Techniques,” Appl. Sci., vol. 11, no. 22, p. 10639, Nov. 2021, doi: 10.3390/app112210639.

[48] K. K. Jena et al., “E-Learning Course Recommender System Using Collaborative Filtering Models,” Electronics, vol. 12, no. 1, p. 157, Dec. 2022, doi: 10.3390/electronics12010157.

[49] Y. Zayed, Y. Salman, and A. Hasasneh, “A Recommendation System for Selecting the Appropriate Undergraduate Program at Higher Education Institutions Using Graduate Student Data,” Appl. Sci., vol. 12, no. 24, p. 12525, Dec. 2022, doi: 10.3390/app122412525.

[50] D. A. Pisner and D. M. Schnyer, “Support vector machine,” in Machine Learning, Elsevier, 2020, pp. 101–121. doi: 10.1016/B978-0-12-815739-8.00006-7.

[51] D. Srivastava and L. Bhambhu, “Data classification using support vector machine,” ***J. Theor. Appl. Inf. Technol.,*** vol. 12, pp. 1–7, Feb. 2010.

[52] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “A comprehensive survey on support vector machine classification: Applications, challenges and trends,” Neurocomputing, vol. 408, pp. 189–215, Sep. 2020, doi: 10.1016/j.neucom.2019.10.118.

[53] C. A. S. Murty and P. H. Rughani, “Dark Web Text Classification by Learning through SVM Optimization,” ***J. Adv. Inf. Technol***., vol. 13, no. 6, 2022, doi: 10.12720/jait.13.6.624-631.

[54] F. Ouatik, M. Erritali, F. Ouatik, and M. Jourhmane, “Students’ Orientation Using Machine Learning and Big Data,” ***Int. J. Online Biomed. Eng. IJOE***, vol. 17, no. 01, p. 111, Jan. 2021, doi: 10.3991/ijoe.v17i01.18037.

[55] B. Gaye, D. Zhang, and A. Wulamu, “Improvement of Support Vector Machine Algorithm in Big Data Background,” Math. Probl. Eng., vol. 2021, p. 5594899, Jun. 2021, doi: 10.1155/2021/5594899.

[56] J. Cao, M. Wang, Y. Li, and Q. Zhang, “Improved support vector machine classification algorithm based on adaptive feature weight updating in the Hadoop cluster environment,” PLOS ONE, vol. 14, no. 4, p. e0215136, Apr. 2019, doi: 10.1371/journal.pone.0215136.

[57] R. Dastres and M. Soori,“Artificial Neural Network Systems,”***Int.J.Imaging Robot***.,vol. 21,pp.13–25,Mar. 2021.

[58] D.-J. Jwo, A. Biswal, and I. A. Mir, “Artificial Neural Networks for Navigation Systems: A Review of Recent Research,” Appl. Sci., vol. 13, no. 7, p. 4475, Mar. 2023, doi: 10.3390/app13074475.

[59] A. Farizawani, M. Puteh, Y. Marina, and A. Rivaie, “A review of artificial neural network learning rule based on multiple variant of conjugate gradient approaches,” ***J. Phys. Conf. Ser***., vol. 1529, no. 2, p. 022040, Apr. 2020, doi: 10.1088/1742-6596/1529/2/022040.

[60] R. Hernández, M. Musso, E. Kyndt, Eduardo Cascallar, and Carlos Felipe, “Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation,” Comput. Educ. Artif. Intell., vol. 2, p. 100018, Jan. 2021, doi: 10.1016/j.caeai.2021.100018.

[61] S.-H. Han, K. W. Kim, S. Kim, and Y. C. Youn, “Artificial Neural Network: Understanding the Basic Concepts without Mathematics,” Dement. Neurocognitive Disord., vol. 17, no. 3, p. 83, 2018, doi: 10.12779/dnd.2018.17.3.83.

[62] S. N. Latifah, R. Andreswari, and M. A. Hasibuan, “Prediction Analysis of Student Specialization Suitability using Artificial Neural Network Algorithm,” in 2019 International Conference on Sustainable Engineering and Creative Computing (ICSECC), Aug. 2019, pp. 355–359. doi: 10.1109/ICSECC.2019.8907173.

[63] S. Gan, S. Shao, L. Chen, L. Yu, and L. Jiang, “Adapting Hidden Naive Bayes for Text Classification,” Mathematics, vol. 9, no. 19, p. 2378, Sep. 2021, doi: 10.3390/math9192378.

[64] F.-J. Yang, "An Implementation of Naive Bayes Classifier," in 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA: IEEE, Dec. 2018, pp. 301–306. doi: 10.1109/CSCI46756.2018.00065.

[65] R. Muthukrishnan and R. Rohini, "LASSO: A feature selection technique in predictive modeling for machine learning," in 2016 IEEE International Conference on Advances in Computer Applications (ICACA), 2016, pp. 18–20. doi: 10.1109/ICACA.2016.7887916.

[66] Z. Jun, "The Development and Application of Support Vector Machine," ***J. Phys. Conf. Ser***., vol. 1748, no. 5, p. 052006, Jan. 2021, doi: 10.1088/1742-6596/1748/5/052006.

[67] A. Kulkarni, D. Chong, and F. A. Batarseh, "Foundations of data imbalance and solutions for a data democracy," in Data Democracy, Elsevier, 2020, pp. 83–106. doi: 10.1016/B978-0-12-818366-3.00005-8.

[68] K. K. Al-jabery, T. Obafemi-Ajayi, G. R. Olbricht, and D. C. Wunsch II, "Data preprocessing," in Computational Learning Approaches to Data Analytics in Biomedical Applications, Elsevier, 2020, pp. 7–27. doi: 10.1016/B978-0-12-814482-4.00002-4.

[69] H. Bon-Gang, "Methodology," in Performance and Improvement of Green Construction Projects, Elsevier, 2018, pp. 15–22. doi: 10.1016/B978-0-12-815483-0.00003-X.

[70] Z. Hazari, G. Potvin, J. D. Cribbs, A. Godwin, T. D. Scott, and L. Klotz, "Interest in STEM is contagious for students in biology, chemistry, and physics classes," Sci. Adv., vol. 3, no. 8, p. e1700046, Aug. 2017, doi: 10.1126/sciadv.1700046.