A proposed Smart System for Determining Appropriate Educational path for Secondary Education Students

نظام ذكي مقترح لتحديد المسار التعليمي المناسب لطلاب المرحلة الثانوية

Mohamed Farh El-galad

Computer Department

Faculty of Specific Education, Damietta University

Prof.Dr. Elsaeed Elsaeed Mohamed Abd El-Razek

Professor of Using Computer in Education, Computer Department

Faculty of Specific Education, Damietta University

Dr. Ahmed Talat Sahlol

Assistant Professor of Using Computer in Education, Computer Department

Faculty of Specific Education, Damietta University

المجلة العلمية لكلية التربية النوعية – جامعة دمياط

عدد (١٠) – ديسمبر ٢٠٢٤

A proposed Smart System for Determining Appropriate Educational path for Secondary Education Students

Abstract

It is difficult (to predict) to determine the student's academic path and the main reason is intervention that is not based on scientific analysis of his previous grades and his academic level in the previous academic stages, for several reasons, including that there is no modern scientific system based on analyzing and monitoring the progress and delay of students in grades and their performance, relying on the desires of The student, regardless of his academic level, human intervention by those closest to him in determining the academic path; The main goal of this research paper is to build an intelligent system capable of analyzing academic grades for various scientific and literary subjects from the preparatory stage until the secondary stage. This paper also focused on how a prediction algorithm can be used to determine the most appropriate academic path for a student. The data of secondary school students was collected, starting from middle school up to high school, and various operations were performed on the data to improve and refine it. Feature engineering operations were also performed, and finally the KNN algorithm was used to classify the students according to one of the two groups, scientific or literary. The proposed machine learning model achieved 92% accuracy in determining the appropriate scientific path for the student, which indicates its superiority compared to other learning models that achieved less accuracy than it.

Keywords:* *KNN - Correlation, Quantization*, *Data Scaling*, *Feature Engineering

1. Introduction

There are many disadvantages affecting the student's scientific future in the event that the exact choice of the appropriate educational path is not made, including the disbursement of a lot of money in lessons for some subjects as his mental abilities do not accommodate such scientific materials.

The student's achievement of unsuitable grades for the college to which he aspires since his scientific path is not commensurate with his mental, psychological and social abilities. Therefore, if we consider and analyze his degrees prior to the school years, we can direct the student to the appropriate scientific path or by dependence does not affect the university education to which he will be enrolled.

Education is the most important sector in the development of a nation. A nation is said to be an advanced nation if the education level of its citizens is good [1].

The quality of education is related to its technological development to keep pace with contemporary education, whose features have been more defined by being more interactive, more individual, and available to everyone and anywhere, relying on the personal computer and information networks that replace the lecture, and the flourishing of distance learning and open learning to replace traditional education and the adoption of curricula on reality life and its economic and social requirements.

Artificial intelligence is a science that seeks to understand the nature of human intelligence through the work of computer programs that are able to simulate human behavior that is characterized by intelligence, through the ability of the program to solve issues and make the necessary decisions, meaning that it is the ability of the machine to imitate and simulate the motor and mental processes of a person, and the way his mind works. In thinking, this is considered an important turning point beyond what is known as "information technology" in which the inferential process is carried out by humans [2].

One of the main focuses of Education 4.0 is to provide students with knowledge on disruptive technologies, such as Machine Learning (ML), as well as the skills to implement this knowledge to solve real-life problems.

Machine Learning is a subset of artificial intelligence that builds a mathematical model based on “training data” to predict an outcome or perform a given task without the machine being explicitly programmed to do it [3].

A number of data mining techniques have already been done on educational data mining to improve the performance of students like Regression, Genetic algorithm, Bays classification, k-means clustering, associate rules, prediction etc [4]. Data mining techniques can be used in educational field to enhance our understanding of learning process to focus on identifying, extracting and evaluating variables related to the learning process of students. Classification is one of the most frequently studied problems by data mining and machine learning researchers [5].

In filed prediction student's performance study Guerrero &Domínguez Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review Predicting students' performance is one of the most important topics for learning contexts such as schools and universities, since it helps to design effective mechanisms that improve academic results and avoid dropout, among other things. These are benefited by the automation of many processes involved in usual students' activities which handle massive volumes of data collected from software tools for technology-enhanced learning. Thus, analyzing and processing these data carefully can give us useful information about the students' knowledge and the relationship between them and the academic tasks. This information is the source that feeds promising algorithms and methods able to predict students' performance. In this study, almost 70 papers were analyzed to show different modern techniques widely a.pplied for predicting students' performance, together with the objectives they must reach in this field.

There are various studies that have created datasets from the results of students in different years in order to predict the performance of schools based on data mining techniques, such as the study of Navamani and Kannammal[6]; Their studies work presents a systematic analysis of various features of the higher grade public school Examination results data in the state of Tamil Nadu, India through different data mining classification algorithms to predict the performance of Schools.

Selim and rezk On predicting school dropouts in Egypt: A machine learning approach[7] A promising technique was proposed in for predicting the risk of dropout at early stages in online courses, where high dropout rate is a serious problem for this kind of courses at university level. This technique is based on a parallel combination of three ML techniques (K-Nearest Neighbor (KNN), RBN, and SVM), which make use of 28 attributes per student.

In ML, classification is a supervised learning technique in which the computer is fed with labeled data and learns from it so that, in the future, it can use this learning to classify new data. The classification may be binary (only two classes) or multi-class (more than two classes). [8] Such as (علمي – ادبي)

2.Methodology

2.1 Dataset Collection

There are still many problems in the educational process, and artificial intelligence had a prominent role in solving them. Therefore, a smart system was created that is able to classify the student scientifically after first secondary school and not rely on traditional methods to solve them according to desires, guesses, and inclinations.

We collect the data from different schools that belongs to the general education system that belongs to different administration in Damietta Governerate. Then we apply different prerocesing on them to be appropriate for machine learning tasks, such as cleaning, removing outliers, scaling.

These are the several grades of our target students, starting from the preparatory stage (the 3rd preparatory class) along with the secondary stage (1nd, 2nd and 3rd secondary class) .

To ensure that we follow the same set of students, we have collected the same stdents names from different educational stages. For example, student 1 was in 3rd preparatory grade in 2016 : 2017, then he promoted to first secondary grade in 2017: 2018, then to second secondary grade in 2018: 2019, then the third secondary grade in 2019: 2020.

It was challenging to track the same set of students along different schools where they transfere from the preparatory grade to the secondary grade, especially when a student move from a city to anther one or even transfer from the general secondary system to the technical education system which has a different branching. Such cases, we have to eliminate them from our pipeline.

Real student degree records were obtained by the Ministry of education, Damietta Branch. Pre-processing is a step that enhances the quality of features and consequently it improves the model's performance.

The use of high-quality data can prevent the consideration of erroneous, missing or redundant information; hence, researchers must collect data from authoritative databases. In 2011, the United States proposed the Materials Genome Initiative for highlighting the importance of massive data in the development of materials science, which strongly encouraged the establishment of a high-quality material database [9].

Table 1 presents the main characteristics of the collected data.

Table 1. Data characteristics

Dataset Characteristics	Multivariate
Attribute Characteristics	Numeric
Associates Tasksarea	Classification
Number of instances (students)	500
Missing Values	Non
Geographic information	Damietta Governorate
Students grade	Primary education instances (students): Secondary education instances (students):
Gender	Male: 217 - Female: 283
Collection Date	1/7/2022

Dataset attributes

Our dataset consists of students from both primary education as well as secondary education, as follows: stage 3 Years preparatory stage 2017/2018 , and the secondary stage as: 1 secondary Years 1^{st} secondary stage2018/2019, Years 2^{st} secondary stage 2019/ 2020 and Years 3^{st} secondary stage: 2020/2021

The features consists of students information beside their grades as seen in Table2.

Table 2: Feature description

Feature vector (29 courses)	arabic3, english3, studies, math3, science3, Computer3, Activity1, Activity2, religious, art_education, Arabic, English, French, algebra, engineering, math, Scientific, literary, Biology, chemistry, History, physics, geography, computer, philosophy, religion, patriotism, Sports_Education, pedagogical
Category	scientific or literal

The students 'grades were extracted from their official grade records, from both, their schools and the educational Directorate. A total number of 500 students records were collected, including the students' grades of the courses from both, the third preparatory grade and the first, second and third secondary grade. While, the total number of all courses from these four grades was 29, as shown in Table (2). For each student record, there is a column for the category, which is scientific or literal (the main two divisions for all students in the Egyptian system in the secondary grade), so it's a binary classification problem.

Real student records were obtained by the Ministry of education, where an a support letter was sent by the faculty of specific education at Damietta University to the Ministry of education. So, our data consists of 500 students with different grades (3^{rd} class of preparatory class, 1^{st} class of class of Secondary, 2^{st} class of class of Secondary, 3^{st} class of class of Secondary), each one includes 29 courses as in table 2, for everyone only one branch is assigned (scientific math, scientific biology and literal).

2.2 Data preprocessing

2.2.1 Data cleaning

Data cleaning is the process of editing, correcting, and structuring data within a data set so that it becomes appropriate for our predication task. This includes removing corrupt or irrelevant data or records with too many missing values. In order to fill the missing values, several replacement strategies have been adopted. First, it can be removed if a feature was not useful. Second, Many student records have been deleted because they have no data in the all classes. Data cleaning is often a tedious process, but it's absolutely essential to get top results and powerful insights from your data.

2.2.2 Data Scaling

In this subsection, we apply some standarization techniques to convert value of features to a specific range. This should speed up the computation time and improve the classification performance because it narrows down the variations along each feature vctor.Data scaling (also referred to as feature normalization or standardization) is concerned with changing the scale of features. When the dataset differs greatly in scale, then a model that is sensitive to the scale of the input features (i.e. linear regression, logistic regression, neural networks) would be affected negatively. Ensuring features are within a similar scale is imperative, whereas, models such as tree-based models (i.e. Decision Trees, Random Forest, Gradient boosting) do not care about scale [8].

One of the most important steps as part of data preprocessing is detecting and treating the outliers as they can negatively affect the statistical analysis and the training process of a machine learning algorithm resulting in lower accuracy.

2.3 Feature Engineering:

Feature engineering is an important but labor-intensive component of machine learning applications [9]. Most machine learning performance is heavily dependent on the representation of the feature vector. As a result, much of the actual effort in deploying machine learning algorithms goes into the design of preprocessing pipelines and data transformations [10].

Also known as feature creation, is the process of constructing new features from existing data to train a machine learning model [11]. This step can be more important than the actual model used because a machine learning algorithm only learns from the data we give it, and creating features that are relevant to a task is absolutely crucial. The following operation have been applied:

Combination: Combining some features and generating new features is one of the useful techniques in machine learning that aims at improving prediction performance by adding more important features. Table 3. (shows multiple combination of features).

Table 3. Combining features that belong the same category with each other

Raw features						Combined (new generated) features
math3	algebra	engineering	-	-	-	**Math**
Scientific3	chemistry	Physics	biology	-	-	**Scientific**
studies3	History	philosophy	geography	-	-	**literary**
math3	algebra	engineering	algebra1	Calculus1	mechanics1	**Math Science**
Biology	Biology2	-	-	-	-	**Biology**

As seen in Table 3, we have generated more five features based on their nature od numeric data. For example, the new added feature; Math is consists of all mathematical related courses, such as math 3 from grade 3rd , algebra from grade 1st and engineering course from grade 1st.

Also, the new added feature; litrary is consists of all literal related courses, such as studies3 from grade 3rd, History from grade 1st, philosophy from grade 1st, and geography from grade 1st.

Quantization: he fundamental idea behind quantization is that if we convert the weights and inputs into integer types, we consume less memory and on certain hardware, the calculations are faster.

Quantization allows the representation of the weights and activations to be as low as 8 bits, or even 1 bit in some cases [12].

Table 4. Transforming continuous numerical values into categories.

Feature	Before		Categories after Quantization			
	Min Value	Max Value	0	1	2	3
Math	49	99	>=٨٠	>=٧٠	>=٦٠	<٦٠
Scientific	59	99	>=٨٠	>=٧٠	>=٦٠	<٦٠
literary	49	100	>=٨٠	>=٧٠	>=٦٠	<٦٠
Math Science	80	158	>=142	>=132	>=122	<٦٠
Biology Science	17	40	>=٣٧	>=٣١	>=٢٥	<٢٤

Correlation : Correlation is a statistical technique that demonstrates how strong is the bonding of a pair of variables. It works for quantifiable data. Correlation in conjunction with linear regression a supervised machine learning (ML) technique.

Correlation has a mathematical formula which when substituted with the values of the variables generates correlation [13].

$$\text{Correlation}=\frac{\sum_{i=1}^{n}(x_i - x^-)(y_i - y^-)}{\sqrt{\sum_{i=1}^{n}(x_i - x^-)^2 \sum_{i=1}^{n}(y_i - y^-)}}$$

Eq. Experssion to evaluate correlation between x and y

Correlation can have a value:

- 0 > correlation <=1 (i.e. positive value), which means it is a perfect positive correlation and means that both variables tend to increase together.
- 0, signifies no correlation meaning the variables x and y have no relation what so ever.

If the correlation between two features is 1.0, that means that they has a perfect relationship. While a 0.9 or 0.8 correlation, means strong relationship, means a strong correlation. While a 0.1 correlation, means a weak correlation.

Table 5 shows correlation between some features.

Table 5. Correlation between some features

	arabic3	**english3**	**studies**	**math3**	**science3**	**Computer3**	**Activity1**	**Activity2**
arabic3		0.630584	0.561675	0.640699	0.525205	0.475887	0.226911	0.228694
english3			0.511014	0.592893	0.516006	0.530615	0.246393	0.20713
studies				0.554762	0.550015	0.43252	0.051379	0.08103
math3					0.560707	0.512997	0.223307	0.213913
science3						0.523814	0.192053	0.191942
Computer3							0.315029	0.33672
Activity1								0.763035

In table 5, we remove the feature, activity 2 because it's highly correlated (.76) with feature activity 1.

3. Classification

K-nearest neighbor algorithm :

This algorithm is used to classify students's academic path and then determine the appropriate path (scientific OR literary). K-nearest neighbor or K-NN algorithm basically creates an imaginary boundary to classify the data. When new data points come in, the algorithm will try to predict that to the nearest of the boundary line [14].

It is one of the simplest and most common classifiers, yet its performance competes with the most complex classifiers in the literature. The core of this classifier depends mainly on measuring the distance or similarity between the tested examples and the training examples. This raises a major question about which distance measures to be used for the KNN classifier among a large number of distance and similarity measures available?

Prasatha et al [15] experimental results show that the performance of KNN classifier depends significantly on the distance used, and the results showed large gaps between the performances of different distances. They found that a recently proposed non-convex distance performed the best when applied on most datasets comparing to the other tested distances. In addition, the performance of the KNN with this top performing distance degraded only about 20% while the noise level reaches 90%, this is true for most of the distances used as well. This means that the KNN classifier using any of the top 10 distances tolerate noise to a certain degree. Moreover, the results show that some distances are less affected by the added noise comparing to other distances.

This classification is done based on the labels of its nearest neighbors. The KNN algorithm has been used on several Industry 4.0 framework applications, such as cybersecurity [18], aircraft's useful life prediction [19], fault classification [20], nephropathy prediction in children [21], intrusion detection systems [22], etc.

Alsariera and etc used KNN for predicting student performance. Especially, academic achievement is one of the metrics used in rating top-quality universities. Despite the large volume of educational data, accurately predicting student performance becomes more challenging [23]. all six articles produced a high level of accuracy in predicting the student's performance. Notably, the highest accuracy was 95.8% [24], and the lowest was 69.

4. Results

4.1 Performance Evaluation

In this paper, we calculated accuracy, precision and F-measure for testing the performance of the proposed approach as seen below.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F-Measure} = 2 \times \frac{(\textbf{Precision} \times \textbf{Recall})}{(\textbf{Precision}+\textbf{Recall})} \quad (4)$$

The researcher conducted many experiments in order to improve the results extracted through the KNN algorithm, including the use of many scaling techniques.

Figure 2 shows a comparison between the chosen scaler in our pipeline; Z-score and other standarization techniques; Standard scaler, Min-Max scaler and L^2 scaler.

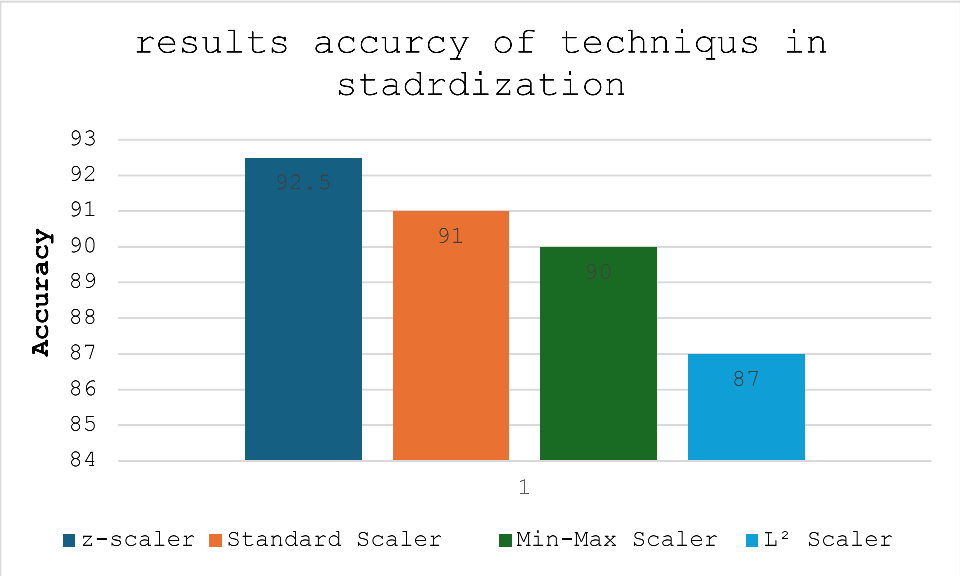

Figure 1. Accuracy results of stadrdization techniques

As shown in figure 1 Z-score achieved the highest classification accuracy compared to the other standardization techniques with a slight advantage than both Standard scaler and Min-Max scaler.

It was the use of a technique Feature engineering A clear effect on the high accuracy of the algorithms used in the system, and we note its strong effect on the accuracy coefficient of the KNN algorithm.

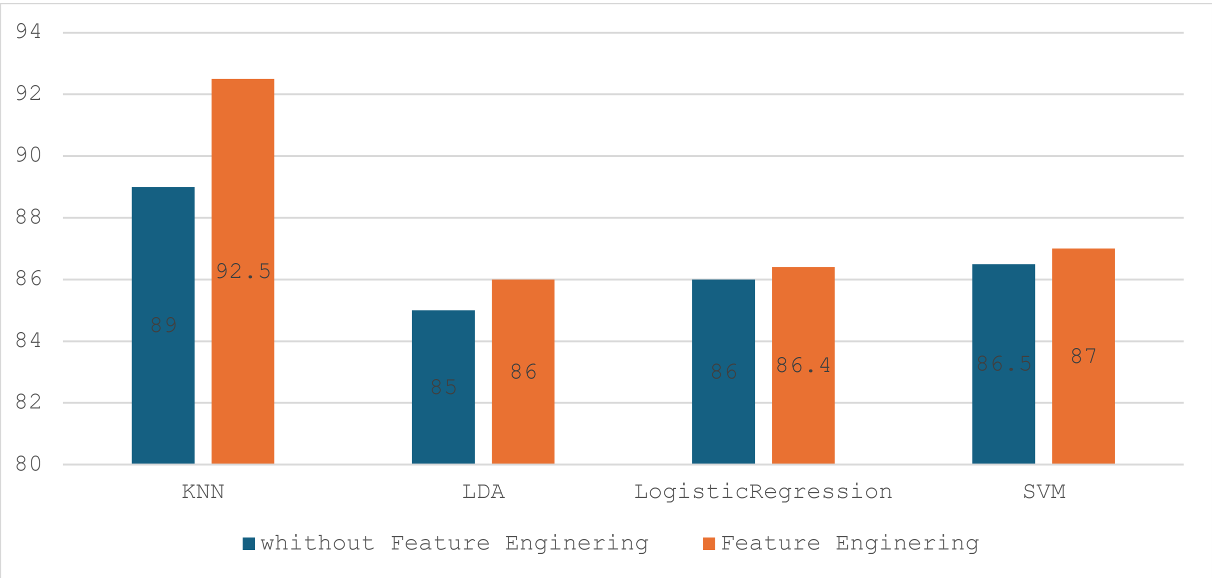

Figure 2. Results with and without feature engineering

As shown in figure 2 feature engineering achieved the highest classification accuracy compared to the other results without feature engineering

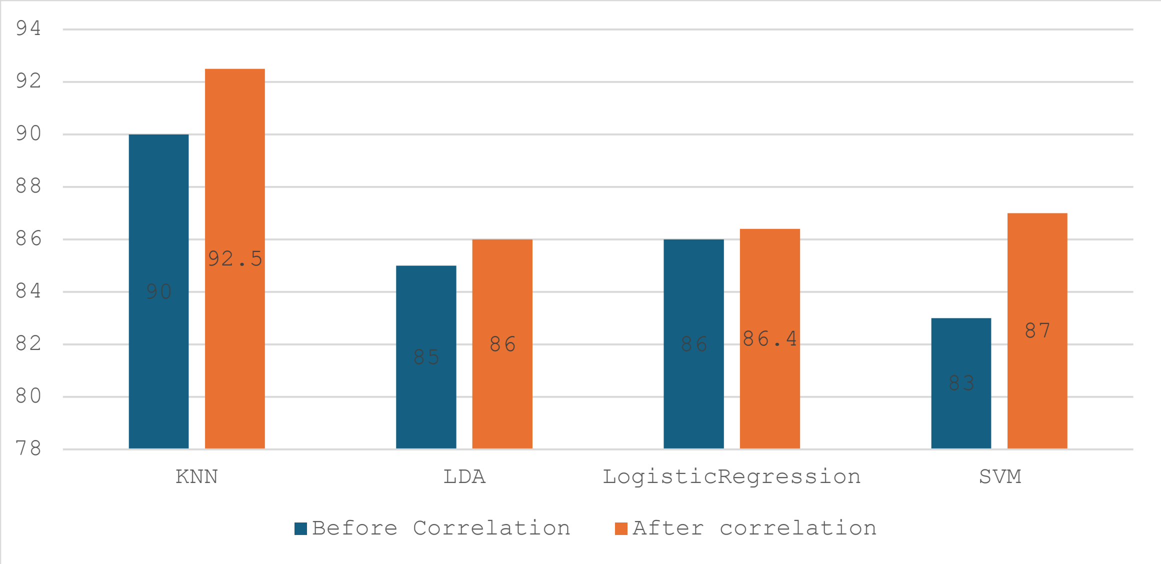

Figure 3. Results before and after correlation

Figure 3 shows the results of machine learning algorithms before and after correlation, where it is clear that correlation effect positively on the classification accuracy for all classifiers and we note its strong effect on the accuracy coefficient of the KNN algorithm.

Table 6. Typical example of A Confusion matrix

		Predicted	
		scientific	**literary**
Actual	**scientific**	TN (61)	FP (3)
	literary	FN (5)	TP (31)

The accuracy can be measured using confusion matrix. Equation below gives the accuracy of the model.

From Table 6, it’s obvious that from 66 sudent chose the sientific branch, 61 of them were crorectly classified by the model, where the rest 5 sudents were missclassified. On the other hand, from 36 sutednts who chose the literary branch, 31 of them were correctly classified by the model, while only 3 students were incrorectly classified as scientific.

Table (7) Detailed performance results for each Class on KNN

	Accuracy	precision	recall	f1-score
Scientific	92.5	0.92	0.95	0.94
Literary	92	0.91	0.86	0.89

In this respect, many experiments have been done using other evaluatison metrics such as precision, recall, and f-messure to determine the efficiency of the classifiers that are used. The resuts that have been obtained show that there is considerable performance variability between the various classifier methode. K nearest neighbor method performs better than other techniqes. Table (8) illustates the experimental results of different classifications performance by these metrics.

Table 8. comparison of different classifications performance using evaluation metrics

Algorithm (Total Instances, 500)	Accuracy	Percetion	Recall	F-Measure
KNN	%92.5	%92	%91	%92
LDA	86%	%86	%85	%81
SVM	%87	%86	%87	%86
LogisticRegression	%86	0.85	0.85	0.85

The experimental results for prediction path scientific future for students' show that the precision, recall and f-measure metric have apercentage of recognition success amounting to 92.5 , 91, 92 respectively using KNN classifier. This means that KNN. is the most reliable classifier and can result in high performance rate in comprise with other classifier's.

From the experimental results of the data sets of students ' grades for previous subjects, we found that the KNN algorithm gave good results with an accuracy rate of 92.5% all the testing samples (25%) of the all data.

4.2 Conclusion

As for the results of the accuracy of the smart system in overcoming the problem of education in the best and appropriate choice of the student for the scientific path compatible with his scientific abilities which provides him and his family with the effort in a path incompatible with his ability according to his previous degrees during the different stages of study. The current study indicates that the use of the smart system has a clear effect in solving one of the problems of education at the accuracy level of Accuracy 92%. The error rate here is 8% as opposed to traditional methods. Traditional methods interfere with the element of stress and subjectivity as well as consuming more time in predicting traditional methods. "What is the accuracy of the results of the proposed smart system in predicting the appropriate scientific course for the student?" The validation of the imposition, which provides for "the use of an intelligent system that is efficient in predicting the student's scientific trajectory"; This demonstrates the intelligent system's clear impact on this forecast; Thus the ability of this intelligent system to learn machine.

Machine learning's effectiveness in predicting student output depends on the good use of algorithms for data. To achieve the best results, it is important to choose the correct machine learning approach for the right data.

The data set of 500 student’s results of a college is considered for the study. The model recommends to the students the best practical path for them according to the grades of the subjects for the previous years.

The data went through several stages, such as: The data went through several stages, such as: Data preprocessing (Data cleaning; Data scaling; Feature Engineering; Correlation) Which had a strong effect on accuracy results.

Based on the confusion matrix of test data, the evaluation metrics like accuracy, accuracy and recall are calculated. It is found that the model, the accuracy is 92.5%, accuracy of 91% and recall value of 86%. The trained machine can Nominating the best educational path for the student with a high level of accuracyو So it can be concluded that this filtering system using machine learning techniques is one of the needs of the hour in this pandemic period.

Researchers recommend the use of artificial intelligence in the electronic coordination processing of university enrolment so that there is a system capable of nominating the appropriate college for the student according to the scientific analysis of the previous school years and not the final collection of secondary certificate.

Reference

[1]A. R., Wahana, A., Maylawati, D. S., & Ramdhani, M. A. (2019, November). Comparison of K-Nearest Neighbour and support vector machine for choosing senior high school. In *Journal of Physics: Conference Series* (Vol. 1280, No. 2, p. 022026). IOP Publishing.

[2]SCIKIT-LEARN, WELCOME TO. "PDF documentation-Scikit-learn."

[3]Zhang, X.D. A Matrix Algebra Approach to Artificial Intelligence; Springer: Berlin/Heidelberg, Germany, 2020.

[4] Selim, K. S., & Rezk, S. S. (2023). On predicting school dropouts in Egypt: A machine learning approach. *Education and Information Technologies*, 1-32.

[5] Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu9, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. *Cancer genomics & proteomics*, *15*(1), 41-51.

[6] Navamani, J. M. A., & Kannammal, A. (2015). Predicting performance of schools by applying data mining techniques on public examination results. Research Journal of Applied Sciences, Engineering and Technology, 9(4), 262-271.

[7] Selim, K. S., & Rezk, S. S. (2023). On predicting school dropouts in Egypt: A machine learning approach. Education and Information Technologies, 1-32.

[8] Ozsahin, D. U., Mustapha, M. T., Mubarak, A. S., Ameen, Z. S., & Uzun, B. (2022, August). Impact of feature scaling on machine learning models for the diagnosis of diabetes. In 2022 International Conference on Artificial Intelligence in Everything (AIE) (pp. 87-94). IEEE.

[9] Sarowar, M. G., Rahman, M., Ali, M. N. Y., & Ripon, S. H. (2019, March). A hybrid machine learning approach for analysis and classification of social network sentiments. In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE.

[10] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798–1828, 2013

[11] Koehrsen, W. (2018). Automated Feature Engineering in Python. *Available on: https://towardsdatascience. com/automated-feature-engineering-in-python-99baf11cc219*.

[12] Courbariaux, M., Hubara, I., Soudry, D., El-Yaniv, R., & Bengio, Y. (2016). Binarized neural networks: Training deep neural networks with weights and activations constrained to+ 1 or-1. *arXiv preprint arXiv:1602.02830*.

[13] K. Q. Weinberger and L. K. Saul, "Distance Metric Learning for Large Margin Nearest Neighbor Classification," Journal of Machine Learning Research, vol. 10, pp. 207-244, 2009.

[14] Islam, M. R., Kamal, A. R. M., Sultana, N., Islam, R., & Moni, M. A. (2018, February). Detecting depression using k-nearest neighbors (knn) classification technique. In *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)* (pp. 1-4). IEEE.

[15] Irfan, M., Nurhidayat, A. R., Wahana, A., Maylawati, D. S., & Ramdhani, M. A. (2019, November). Comparison of K-Nearest Neighbour and support vector machine for choosing senior high school. In *Journal of Physics: Conference Series* (Vol. 1280, No. 2, p. 022026). IOP Publishing.

[18] Romeo, L.; Loncarski, J.; Paolanti, M.; Bocchini, G.; Mancini, A.; Frontoni, E. Machine learning-based design support system for the prediction of heterogeneous machine parameters in industry 4.0. Expert Syst. Appl. 2020, 140, 112869. [CrossRef]

[19] Taha, H.A.; Sakr, A.H.; Yacout, S. Aircraft Engine Remaining Useful Life Prediction Framework for Industry 4.0. In Proceedings of the 4th North America conference on Industrial Engineering and Operations Management, Toronto, ON, Canada, 23–25 October 2019.

[20] Zhou, C.; Tham, C.K. Graphel: A graph-based ensemble learning method for distributed diagnostics and prognostics in the industrial internet of things. In Proceedings of the 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS), Singapore, 11–13 December 2018; pp. 903–909.

[21] Zhang, P.; Wang, R.; Shi, N. IgA Nephropathy Prediction in Children with Machine Learning Algorithms. Future Internet 2020, 12, 230.

[22] 39. Thapa, N.; Liu, Z.; Kc, D.B.; Gokaraju, B.; Roy, K. Comparison of machine learning and deep learning models for network intrusion detection systems. Future Internet 2020, 12, 167. [CrossRef]

[23] Alsariera, Y. A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., & Ali, N. A. (2022). Assessment and evaluation of different machine learning algorithms for predicting student performance. *Computational Intelligence and Neuroscience*, *2022*.

[24] D. Kurniadi, E. Abdurachman, H. L. H. S. Warnars, and W. Suparta, "*e prediction of scholarship recipients in higher education using k-Nearest neighbor algorithm," IOP Conference Series: Materials Science and Engineering, vol. 434, no. 1, Article ID 012039, 2018.

نظام ذكي مقترح لتحديد المسار التعليمي المناسب لطلاب المرحلة الثانوية

من الصعب (التنبؤ) تحديد المسار العلمي للطالب والسبب الرئيسي التدخل الغير قائم على التحليل العلمي للدرجات السابقة له ومستواه العلمي في المراحل العلمية السابقة، وذلك لعدة أسباب منها لا يوجد نظام علمي حديث قائم على تحليل ورصد تقدم وتأخر الطلاب في الدرجات وأدائهم ، الاعتماد على رغبات الطالب دون النظر لمستواه العلمي، التدخل البشري من قبل الأقربون في تحديد المسار العلمي؛ الهدف الرئيسي لهذه الورقة البحثية هو بناء نظام ذكى قادر على تحليل الدرجات العلمية للمواد العلمية والأدبية المختلفة من المرحلة الإعدادية حتى المرحلة الثانوية. هذه الورقة أيضًا ركزت على كيفية استخدام إحدى خوارزمية التنبؤ لتحديد المسار العلمي الأنسب للطالب. تم تجميع بيانات طلاب الشهادة الثانوية بدء من المرحلة الإعدادية وصولًا للثانوية وتم إجراء عمليات مختلفة على البيانات لتحسينها وتنقيحها وأيضًا تم إجراء عمليات هندسة السمات ثم في النهاية استخدمت خوارزمية KNN لتصنيف الطلاب وفقًا لإحدى المجموعتين، العلمية أو الأدبية. حقق نموذج تعلم الآلة المقترح دقة ٩٢% في تحديد المسار العلمي المناسب للطالب مما يدل على تفوقه مقارنة بنماذج تعلم إلى أخرى حققت دقة أقل منه.

الكلمات المفتاحية: نظام ذكي – المسار التعليمي – لطلاب المرحلة الثانوية