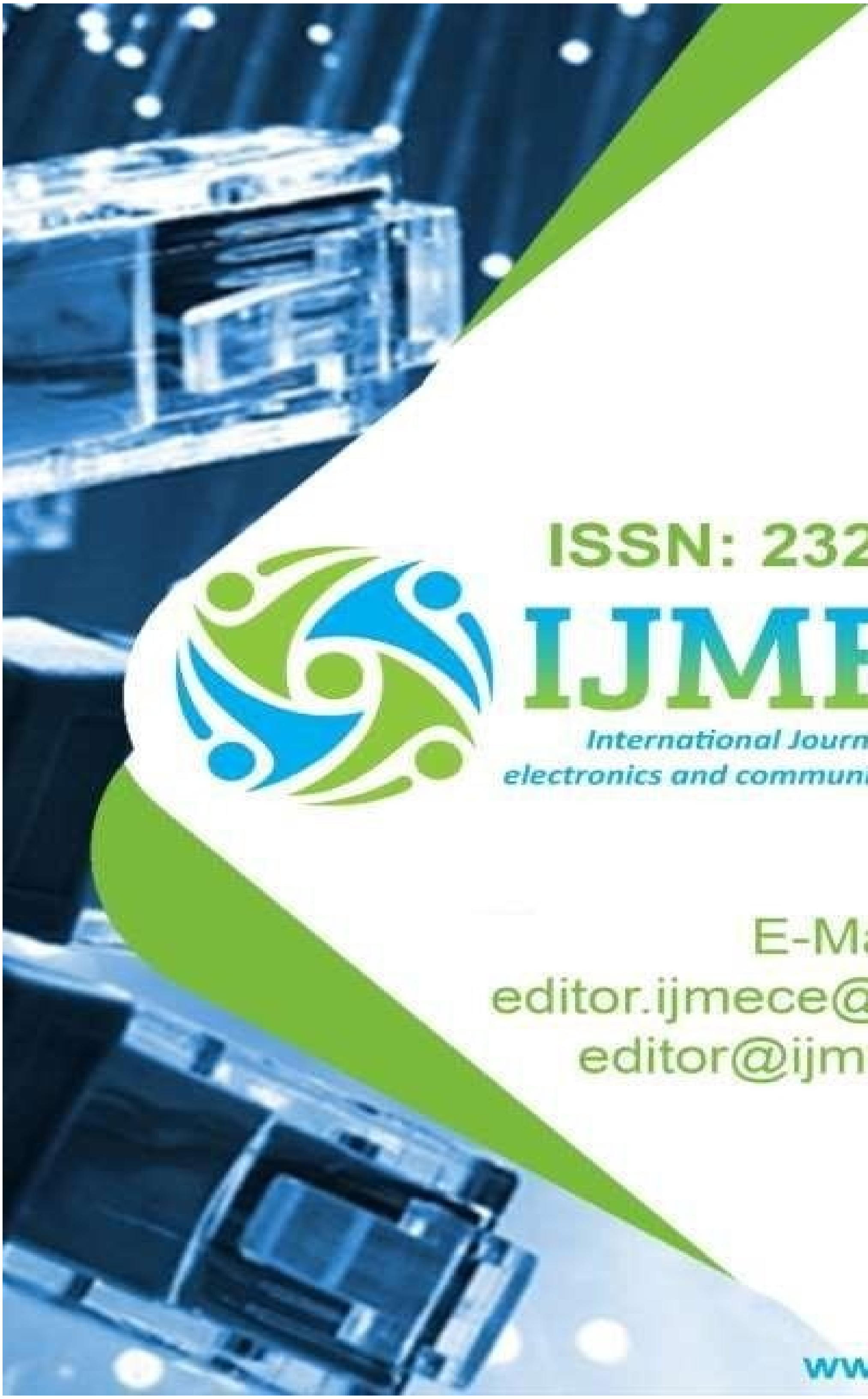


ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

Creating a System to Identify Students at Risk in Order to Assist with Educational Planning

[1]Ms. Y.Bindu Rajasri, [2]Mattaparthi Hara Veera Sai Teja,

[1]Assistant Professor,Dept.of MCA, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

[2]Students,Dept.of MCA, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

ABSTRACT: Educational data mining (EDM) emerged as a result of the profound influence that advances in data analysis and intelligent systems have had on the educational system. If you're looking to analyze student performance or identify at-risk pupils, the Early Warning System (EWS) is a lifesaver. Our study focuses on creating a system that can detect when students are about to drop out of school by considering several educational, socio-cultural, and structural elements. In an effort to be as thorough and precise as possible when selecting dropout indicators, we have developed a unique database just for this purpose. With a training set accuracy of over 99.5% and a test set accuracy of over 99.3%, our model did very well, especially when using the K-Nearest Neighbor (KNN) method. We present a Django app we built specifically for this purpose, which visualizes the findings, and demonstrate how this may be used to educational planning.

INDEXTERMS: Earlywarningsystem,machinelearning,KNN,educationalplanning,dropout.

INTRODUCTION:

The emergence of intelligent systems, particularly recommendation and predictive systems, has altered the traditional trajectory of information technology practice. These systems have discovered even more opportunity to thrive and accomplish the most astounding achievements with the advent of Big Data [1] and the subsequent explosion of data. One of the most well-known forms of intelligent systems, early warning systems (EWS) have profited greatly from the recent advances in computing methodologies and technologies, along with the improvement of hardware infrastructures.

In order to aid in decision-making, the EWS is a predictive system that uses studied data to provide a proactive picture of the future. EWS are ubiquitous; their job is to spot out-of-the-ordinary occurrences in actual systems and alert decision-makers to the gravity of the issue, allowing them to prepare for potential interventions and either fix the problems at hand or mitigate their negative impacts. The following active keywords may be used to describe an EWS: collect, analyze, detect, prevent, alert, and notify [2], [3]. Each letter stands for a critical phase in an EWS, and together they form its action model, which is a series of interconnected processes. Collecting data in real-time or near-real-time and continuously monitoring important indicators is the first stage. Processing and analysis of the acquired data is the next phase. Finding early signs, trends, or outliers in the data is what this approach is all about. To spot possible outliers in this research, a number of methods may be used, including sophisticated algorithms, statistical models, or AI. Finally, in the "Alert and Notify" phase, the system promptly notifies the relevant parties whenever it identifies a possible danger or anomaly. Next, in the fourth phase called "Risk assessment," experts and managers check how serious and dependable the warning was. In order to comprehend the nature of the risk, its probable outcomes, and ways to lessen its effect, they review the available data and information. An important part of managing and responding to risks is the communication and dissemination process. Sharing the appropriate information with stakeholders, decision-makers, and the public is vital after assessing and verifying a risk. Responding and taking action constitute the last stage of the procedure. Lastly, when the early warning system has supplied the required data, the right actions are taken to lessen danger, avert disasters, or lessen their negative impacts. Due to the vital nature of EWSs and the fact that their outputs might, in certain situations, prevent catastrophes from happening, both now and in the future, these processes constitute an iterative process in which we always strive for system perfection. For that reason, we never rest until we achieve perfection.

EMISANDEDM

TheEducationTo aid in educational decision-making, the Education Management Information System (EMIS) facilitates the gathering and storage of three distinct kinds of educational data: fundamental, business, and statistical. After that, we'll use the right technology to examine these data sets. In the end, we want to provide decision-makers some visual reports or findings that cover all the bases in terms of education.

The term "Educational Data Mining" (EDM) encompasses a wide variety of techniques used to mine educational databases for useful information [4]. Statistics, data mining, and ML-related methods fall under this category. At its core, the EMIS is EDM [5, 6], the machine that processes data, creates data models and profiles of education stakeholders, and is aware of all the linkages and interconnections among this data.

Data is the lifeblood of the school management information system, making it undeniably a data-driven system. Future strategies and decision-making in education may be informed by analyzing this data using EDM approaches and their outcomes.

The EWS is an essential component of the Learning Management System (LMS) or the EMIS system as a whole because of the role it plays in processing and converting educational data for targeted applications. With the help of EDM, these aims have been varied, which means that EWS alters both the facet and the output. Analyzing students' progress and grades in the subjects or courses they have taken is one of the key themes covered by EDM, which is students' performance analysis [6]. The study of learner behavior is another theme. This can be seen through the lens of the learner's experience in their school or academic environment, with all its variables operating independently, or through the lens of the learner's interaction with course and curriculum content. Dropout prevention is another focus of the EDM [4]. This issue has received a lot of attention from educational institutions because to the damage it does to students' futures and their ability to study. The creation of EWS [8] is a natural outgrowth of research into educational phenomena; its purpose is to forestall the onset of such phenomena and allow decision-makers the strategic room they need to tackle their root causes before they spiral out of control. Given the increasing need for better data visibility and situation-specific advice among educational field actors, it has also allowed advanced research into decision-support systems [9]. To wrap things off, we'd want to talk about something a little more universal: how effective our educational systems are [10]. While this is more appropriately addressed by formal national or international organizations, EDM methods may be quite useful for comparing actual outcomes with planned educational outcomes. We could not discover any discussion of this strategic level of educational planning in the subject of EWS development research in education, therefore we have given it a lot of weight in our work. The reasoning behind this is that in order to understand any issue related to education, we must put it in a broader global perspective and address it via strategic educational planning.

EDUCATIONALPLANNING

EIt is widely acknowledged that education is fundamentally a social issue due to its impact on individuals. Furthermore, the literacy rate and the general education level of a country's citizens are clear indicators of its degree of development. The consequences of education don't manifest in behavior and accomplishment right once, which is why educational planning is essential for reaching any educational objective.

In order to accomplish society's educational goals, educational planning is the methodical and logical process of developing and implementing an educational plan that takes into consideration all available resources, whether they be human, financial, or material. Educators have the responsibility of ensuring that all students have access to a high-quality education in a fair and efficient manner, as the right to education is a fundamental human right that is upheld by all international treaties.

Because it shows students fleeing from the education system, particularly before they have finished the mandatory level of instruction, dropping out of school is a stain on the implementation of education policies. However, this also means less resources will be available to keep pupils in school, and the worst part is that they won't be prepared for the real world when they graduate. That is why it is crucial to have educational plans that are both adaptable and modern, making use of new methods to ensure that they are both thorough and efficient.

Ourpaperisorganizedasfollows. Our literature assessment of related programs that have created

EWS or EDM solutions for dropping out of school is presented in Section II. The components of our EWS construction strategy are detailed in Section III. Also, in Section IV, we go into the many approaches that we used to develop our solution proposal. Afterwards, in Section V, we detail the outcomes of our approach to the issue of student dropouts. Section VI concludes with a summary of our concepts and a discussion of our future plans.

RELATEDWORK

Particularly in less developed nations, EWS is not often used in the classroom. However, with all the benefits it provides, a predictive system is an absolute must. Several organizations and efforts have been attempting to implement EWS in the educational system, and our paper's contribution interacts with one of them. And EWS is an area that has spurred a lot of activity in terms of output value recently. Educators and researchers have been working hard on several fronts to improve classroom instruction and reduce student dropout rates [11]. The analytical logic, the area of practice, and the system's aim have all influenced the ways in which the different EWSs have operated. Some possible ways to categorize these elements are as follows: • Factors related to the school [5, 8, 12, 13, 14, 15, 16, 17],

in references [18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29]: Show the many forms that influence how well pupils do in the classroom. First, we look at the student's educational history, which includes details like where the student is from and whether or not they had pre-school. Where did they get their education? Was it official, informal, or unconventional? The opposite side of the coin is the student's academic record from the years of schooling leading up to the study (including course and test grades, overall averages, and final scores). When examined cumulatively over time and, ideally, within the context of a cohort, these elements become even more significant in determining a student's legacy. Researchers sometimes have to make assumptions based on a school year's worth of data since there just isn't enough information to do a time-series analysis. In any event, school-related variables do not provide a comprehensive investigation of educational phenomena; at most, they permit a statistical evaluation of academic achievement.

The role of people All the things that make the student unique as an individual, including his gender, age, ethnicity, and place of birth [5, 16, 23, 30, 31, 32]. At this level, we also discuss the habits and characteristics of students' behavior, particularly those that impact their ability to learn and function academically, such as tardiness, lack of focus, etc. A student's emotional and physical well-being are also considered, as these factors may have a major impact on how well they study. When investigating the nature and personality of the learner, these considerations are crucial. Either empirical investigations or one-on-one examination of students may provide behavioural data. However, in developing nations, medical data is scarce, which might lead to the neglect of a crucial factor that could explain why students are struggling or losing what they've learned. When trying to forecast educational events, human indicators are inadequate since people do not exist in a vacuum but rather within a larger community and environment.

Environmental considerations [15], [33]:` We talk about a wealthy school in a resource-rich environment and an impoverished school in an underprivileged setting in education sociology because the environment unquestionably has a very big influence on the person. At this level, we also classify schools by kind; for example, private schools often outperform public schools in terms of academic outcomes, mostly because of resource differences. The internal state of the family, or household status in the general population census, is another crucial component of these characteristics. The latter sheds information on markers of the family's socioeconomic standing, educational attainment, cultural norms, and goals for the future (parents, siblings, and cousins). Last but not least, the internal climate of the school is significant in this regard, including the management model, the facilities, the condition of the teacher-learner connection, and the maintenance of internal peace and nonviolence. The influence of the school environment and its actors on students is emphasized by these aspects. However, we must exercise caution when considering the factors to be employed and how the environment is perceived; for instance, we must ask ourselves how much of an influence the family has in comparison to the neighborhood or the school. Media and social network effects, the influence of upper-class students on freshmen, and other environmental variables may also play a significant role that we were unable to measure.

• Noteworthy elements [6, 18, 20, 22, 24, 26, 32, 34, 35]: Exceptional circumstances and natural catastrophes that disrupt schooling fall under this category. This includes things like wars, natural

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

disasters, and human disputes. The coronavirus pandemic is one of the most remarkable events of the last three years. Schools throughout the globe shuttered their doors as a result of the epidemic, and new forms of education, including as distant learning—in which e-learning and hybrid approaches thrived—took their place. Only with extraordinary eyesight can one study these situations, which often have a drastic and sudden impact. However, while creating a generic model for sustainability, they are ignored.

It is critical to include all facets of a learner's circumstance and activity while conducting a sustained analysis to evaluate their performance [11]. We are happy to focus on a subset of thematic indicators and ignore the others because this is so seldom possible. For the initiatives mentioned above, this is the most important point to note. The great majority are employed by institutions of higher learning, with a considerable number specializing in online and hybrid educational environments. As a result, empirical studies of students' varied learning environments are underrepresented, particularly those focusing on the primary and secondary school years, when the dropout rate is greatest and a worrying large-scale indication of socioeconomic status and literacy is evident. We must approach these levels of study with a rigorous and thorough scientific perspective, since they constitute the cognitive and formative foundation of students' talents, and give them our full attention in our research efforts.

One more thing: all the papers use different approaches and cutting-edge tech to get the best results. However, the models are still not precise enough, maybe because they're used on generic data sets and in countries or situations that aren't related to the researchers. In addition to a multidisciplinary field that incorporates disciplines such as demographics, economics, statistics, human sciences, psychology, management, and others, one must possess understanding of educational planning and the area of education in order to choose appropriate research parameters. The number 36. Consequently, this paper's attempt to build an EWS to detect the phenomena of dropouts begins with a thorough understanding of the subject of educational planning, which necessitates an all-encompassing perspective of the educational process and its many stakeholders and interdependencies. The next step is to master the inputs required for each analysis so that they can accomplish their goals. Then, you'll need to be able to explain the predictions that come out of them and make the right recommendations so that you can make decisions that involve structural, organizational, or political actions rather than just changes to teaching methods, curricula, or the learning experience.

MATERIALS

INFORMATIONCOLLECTION

Using Morocco as a case study, we examine the country's educational system in relation to those in South Asian emerging nations, particularly those in Africa.

At the beginning of the school year, there is data that includes all the students who are expected to continue their studies this year. In the middle of the year, there is data that managers use to verify the accuracy of the data, particularly when it comes to the students' situations. This data provides a snapshot of two different times of the year. An improvement in the data kept in the system, as well as their compatibility with the actual situation, is the goal of using the census reference date. Both for taking stock of the present and for meticulously preparing for the next school year, this is invaluable. Immediately after the posting of final test results is the second information-gathering moment. This gives us a reasonable notion of the students' academic performance and lets us investigate their academic accomplishments on the basis at the beginning of the school year.

INFORMATION SET

This project's database was custom-built by merging data from many operating systems, as seen in See Figure 1. The dropout phenomenon may be better understood and analyzed with the help of the information provided by each system:

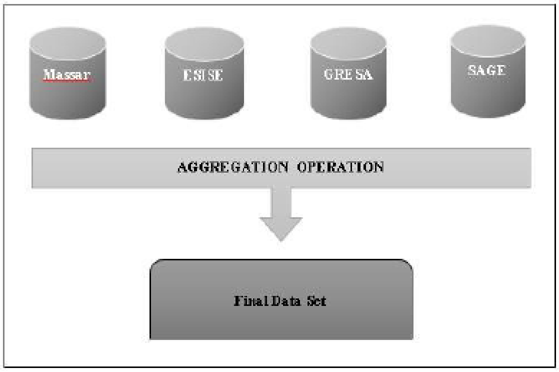


FIGURE 1. Aggregation of data from various operational educational information systems.

During this school year, Massar gives you information on the student, including his or her attendance and grades, as well as other details about his or her past. Oh, and let's not forget

IJMECE
International Journal of modern electronics and communication engineering
ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

about the grades. Use ESISE to get information on educational delays and other metrics tracked across a student's academic career. GRESA: This helps us understand the socioeconomic status of the students in question and identifies which communities get social support for students. SAGE is an exam management system that helps us gauge how well students do on standardized tests in their last years of school. Itwasdecidedtoconfinethisresearchtotheprimary

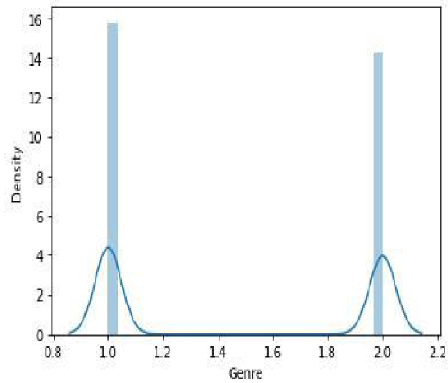


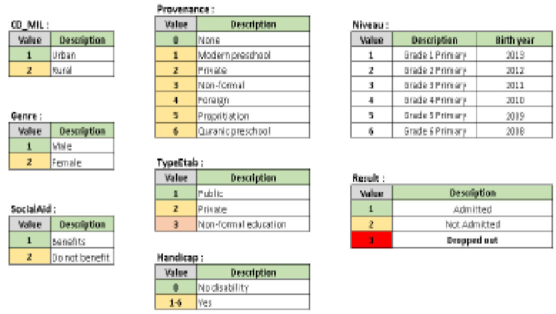
FIGURE 2. The density of male and female students in the experimental data set.

cycle, and potentially to extend the conclusions to the subsequent school cycles. There are 12,5354 students in the study, with 5,9384 of them being female. See Table 1 for statistics on the database's partitioning classes, and see Fig. 2 for a visual representation of the gender distribution. The binary attributes will be explored first. The "CD MIL" feature determines whether the student is from a rural or urban location. The term "genre" is used to de-sex the group under study. Finally, the "SocialAid" function shows which students are helped by social assistance. In practice, this assistance is most visibly manifested via the 'Tyssir' program for tutoring, which distributes grants to families with school-aged children.

TABLE 1. Distribution of data set classes.

Class	Code	Total
Admitted	1	98543
Not Admitted	2	7838
Dropped	3	18973

When analyzing the subjectofourmodel student's state, multi-criteria features are a great way to add more dimensions. To start, the "Provenance" feature may take on a value between 0 and 6, with 0 indicating that the pupil has never benefitted from any kind of pre-school education, whether it be formal or non-formal, contemporary or traditional, national or international, or any other combination.



CD_MIL :

Value	Description
1	Urban
2	Rural

Genre :

Value	Description
1	Male
2	Female

SocialAid :

Value	Description
1	Benefits
2	Do not benefit

Provenance :

Value	Description
0	None
1	Modern preschool
2	Private
3	Non-formal
4	Foreign
5	Propitiation
6	Quranic preschool

TypeEtab :

Value	Description
1	Public
2	Private
3	Non-formal education

Handicap :

Value	Description
0	No disability
1-6	Yes

Niveau :

Value	Description	Birth year
1	Grade 1 Primary	2013
2	Grade 2 Primary	2012
3	Grade 3 Primary	2011
4	Grade 4 Primary	2010
5	Grade 5 Primary	2009
6	Grade 6 Primary	2008

Result :

Value	Description
1	Admitted
2	Not Admitted
3	Dropped out

FIGURE 3. The description of features values.

Learners' access and capacity to succeed in the main cycle and prevent early drop-out are significantly impacted by this factor. Formal, informal, or public education are all represented by the "TypeEtab" attribute. The next step is to identify the six different kinds of disabilities, and the "Disability" function will reveal whether or not a pupil has one. Especially in an unequal educational system, disability is a big element influencing children's schooling. We emphasize the projected year of birth of the student at each school level for use in the academic delay calculation; the values of the "Niveau" characteristic are the primary cycle grades. Last but not least, the "Result" target property, as shown in Table 1.

METHODS

The prediction mechanism is the essential component of a generic data processing model that must be adopted by the implementation logic of an EWS. However, extensive pre-processing is necessary due to the difficulties of collecting data from several sources. In addition to imputation of empty and data extraction, the data arrives lacking and inconsistent, necessitating a solidification by join in accordance with the student code.

TABLE 2. List of transformed features.

Feature	Type	Encoding
Type Etab	Number	Integer 64
Genre	Number	Integer 64
Niveau	Number	Integer 64
RetardSco	Number	Integer 64
Provenance	Number	Integer 64
Handicap	Number	Integer 64
SocialAid	Number	Integer 64
CD_MIL	Number	Integer 64
Moy	Number	Float 64
Result	Number	Integer 64

It was quite difficult to comprehend and use the data collected from the many operating systems.

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

This is particularly true given the design of the systems in question, which consists of a school directory management system (GRESA), an examination management system (SAGE), a census system (ESISE), and a school management system (Massar). There are additional systems connected to various education management professions, such as the human resources management system (MasiRH), the school mapping system (CarteSco), RAED, PSP, and others, but this list is restricted to the ones we were able to employ in this project.

$$\text{Academic Delay} = IF\left(Niveau == 1; \left(2014 - Year_{ofBirth}\right); IF\left(Niveau == 2; \left(2013 - Year_{ofBirth}\right); IF\left(Niveau == 3; \left(2012 - Year_{ofBirth}\right); IF\left(Niveau == 4; \left(2011 - Year_{ofBirth}\right); IF\left(Niveau == 5; \left(2010 - Year_{ofBirth}\right); IF\left(Niveau == 6; \left(2009 - Year_{ofBirth}\right)\right)\right)\right)\right)\right)\right)$$

Despite the Moroccan Ministry of Education's best efforts, the multi-system situation remains, and the huge limits it imposes make it impossible to merge all of these operational information systems into a single integrated system.

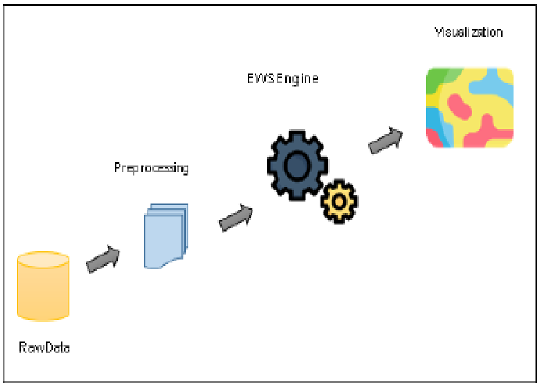


FIGURE 4. The proposed model of the predictive data processing system.

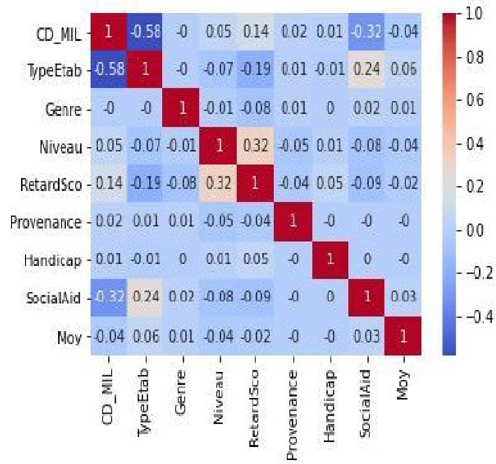


FIGURE 5. Correlation matrix of the experimental data set.

Initial Steps

It took a lot of pre-processing operations, but the database that resulted from cross-referencing all of those information system databases is quite vast. We dealt with the student's national code one second and their student identity the next, thus it was challenging to discover a unique identifier to use as a natural link across all the databases. Attributes that were meant to be treated as integers, among others, need type conversion. Below is a list of the qualities that are going to be transformed:

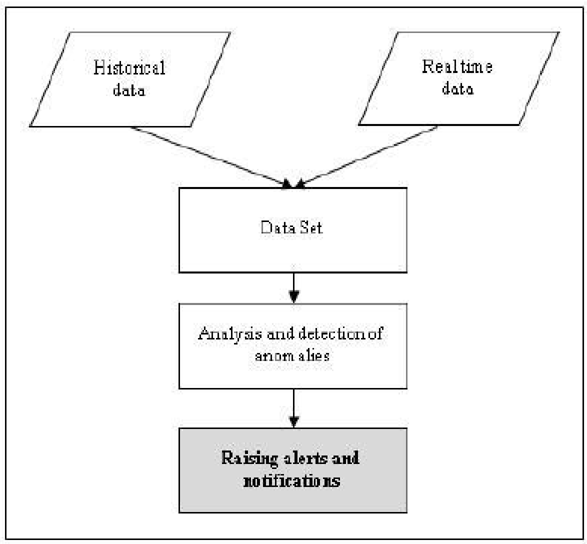


FIGURE 6. The EWS flow chart.

When feasible, we've translated the values of all attributes to binary numbers (1 and 2); this includes features like TypeEtab, Genre, Handicap, SocialAid, and CD MIL. The "Moy" attribute is a continuous variable, and the other

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

features "Niveau" and "Provenance" are also kept unchanged as they include several potential values. Because some systems code men as 1 and females as 2, while others code them the other way, there has also been regulation and coding/decoding operation involving the "Genre" characteristic.

1)SVM

One of the many algorithms that make up supervised machine learning is the support vector machine (SVM). Classification and regression are the bread and butter of this algorithm's design. Its efficacy becomes more apparent when handling complicated datasets with many attributes. Finding a hyperplane that optimally divides different classes of data points is the basic purpose of support vector machines (SVM). This hyperplane acts as a decision boundary.

For a linearly separable case:

$$\omega.x + b = 0$$

where:

- w is the weight vector perpendicular to the hyperplane.
- x is the input feature vector.
- b is the bias term (also known as the intercept).

Here is a representation of the hyperplane equation:
Their decision to drop out of school at any time during the school year may be explained for a linearly separable scenario. Figure 6 shows the sequential organization of the analytical system that is fundamental to the early warning system.

$$\omega.x + b - \varepsilon = 0$$

For the most part, we relied on classification-focused machine learning analytical techniques. This study relies on the most well-known and effective methods: SVM, Random Forest, SGD, and KNN. In order to employ the best possible version, we have attempted to calibrate the internal parameters of the algorithm.

$$\text{Random Forest} = \sum \text{Decision Trees}$$

The K-Nearest Neighbours (KNN) method is a straightforward and user-friendly machine learning tool for regression and classification jobs. As an example of instance-based learning, it takes the feature space's k closest data points as an average or majority class and uses it to generate predictions. When calculating KNN algorithms, several distance measures are used. The two most popular methods involve adding up the absolute differences between the points' coordinates. When two places are measured along the gridlines, the result is the Manhattan distance, which is also called the taxibor city block distance.

$$\theta_{t+1} = \theta_t - \alpha.\nabla L\,(\theta_t; x_i, y_i)$$

where:

- θ_t is the vector of model parameters at iteration t.
- α is the learning rate, a positive scalar that determines the step size of the update.
- $\nabla L(\theta_t ;x_i, y_i)$ is the gradient of the loss function L with respect to the model parameters θ_t, evaluated on the data point (x_i, y_i).

The anticipated outcome of this processing is a predictive model that may be used for future long-term planning operations. This will make the issue of student dropouts more visible and provide educational decision-makers with more room to respond.

$$\text{Euclidean Distance}\,(x, y) = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}$$

We will use a set of the most effective measures to be confident in the reusability of our model, and the EWS will also assist a process of data and model review. This strategy runs the risk of making us make the incorrect action and choice for the scenario at hand if our predictions are inaccurate about the future of one or more kids.

- $\textit{Sensitivity} = TPR = \dfrac{\textit{True Positives}}{\textit{True Positives} + \textit{False Negatives}}$
- $\textit{Specificity} = FPR = \dfrac{\textit{False Positives}}{\textit{False Positives} + \textit{True Negatives}}$

An assessment

How well a model performs depends on how well it can mimic its counterparts in the original test dataset (Y) in order to improve its accuracy in predicting outputs (yˆ). While doing database analysis, the other side of the operation is to keep the margin of error to a minimum. Identifying which parameters are damaging to the model's performance may be done by studying the trend in the amount of error. This allows for direct regulatory action. Here are the metrics that were used:

IMAGE CREATION

Data visualization is a crucial part of every data processing operation. This allows decision-makers to reflect on the measures required to rectify the difficulties faced and provides a consistent image of the analysis process's conclusions. It also provides a framework for further exploration of the scenario under investigation.

TABLE 3. Performance measurements for the algorithms used.

	TRAINING SET			TEST SET		
	MAE	**RMSE**	**R2**	**MAE**	**RMSE**	**R2**
Random Forest	0.0057	0.0606	0.9931	0.0090	0.0868	0.9858
SVM	0.0095	0.1068	0.9913	0.0090	0.1051	0.9919
SGD	0.0123	0.1193	0.9885	0.0120	0.1183	0.9889
KNN	0.0056	0.0863	**0.9953**	0.0076	0.0988	**0.9933**

We have created an application using an MVC architecture as the end result of our labor. It was built utilizing the Bootstrap MDB package and the Django 4.2 Framework. The distribution environment of Anaconda 1.10 with the top ning. First, the application will have to gather data, which we will first provide to the system as a global CSV file for direct processing. In contrast, system-level analysis and visualization will be used for the data set.
We have settled on Tables, Maps, and Charts as our primary means of displaying results for the time being. Overall student information and alerts of those likely to dropout are shown in tables. Using themed maps and graphs, we first identify the communes hit hard by the dropout phenomena, and then show how the phenomenon varies throughout these communes using graphs.

RESULTSANDDISCUSSION

We computed the coefficient of determination R2 to gauge our model's accuracy, but we also tried out a number of other approaches and techniques to get at the most crucial metric in this study. To create the loss curve and choose the most accurate model, we further used the MAE and RMSE metrics.

Model for EWS's Line.

Due to the categorization aspect of the issue, we choose to investigate the most well-known algorithms in this area, citing: SVM, Random Forest, SGD, and KNN. You may confidently reuse the model since the results, as shown in Table 3, are quite acceptable. With an R2 of more than 99.5% on the training set and more than 99.3% on the test set, the KNN algorithm ultimately achieved the greatest results, recording the lowest loss (MAE&RMSE) and generating the highest accuracy score (R2).

Important Features

TABLE 4. The weight of each feature in the target prediction.

Feature	Weight
Moy	0.9962324622
RetardSco	0.0010250153
Genre	0.0008447858
Provenance	0.0007557517
Niveau	0.0006246890
SocialAid	0.0002653846
CD_MIL	0.0001764032
TypeEtab	0.0000722366
Handicap	0.0000032718

A variety of characteristics allowed for a specific categorization of the training set data and served as the basis for the model that was eventually constructed. With respect to the selection of model output (Target label), Table 4 displays the weight of each characteristic according to the significance of its involvement.

$$\begin{aligned}
&\textit{Dropped out} \\
&\quad = \{CD_MIL == 2; TypeEtab == 1; \\
&\qquad Genre == 2; Niveau == 4; \\
&\qquad RetardSco == 2; Provenance == 1; \\
&\qquad Handicap == 0; SocialAid == 2; Moyenne == 0\}
\end{aligned}$$

Let's assume that the grade point average is the most crucial part of the label description. This number provides a composite view of a student's academic performance by combining their test scores with their grades from all of their classes, particularly at the sixth-grade certification level. Academic delay, which is another way of saying grade repetition, is the second most important feature since it affects the student's future academic advancement.

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

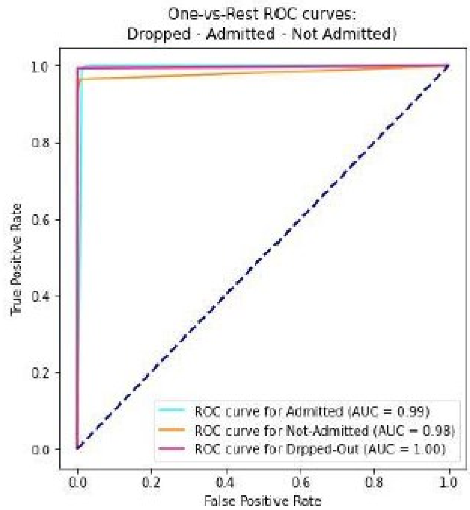


FIGURE 7. ROC Curve of Label class modalities.

Because gender is still one of the first elements to influence school dropout in poor nations, the model also places a high value on the "Genre" feature. Because of this, the qualities "Provenance" and "Niveau" are equally important. Part of a student's educational history includes their background before attending elementary school, which may influence their decision to continue their studies.

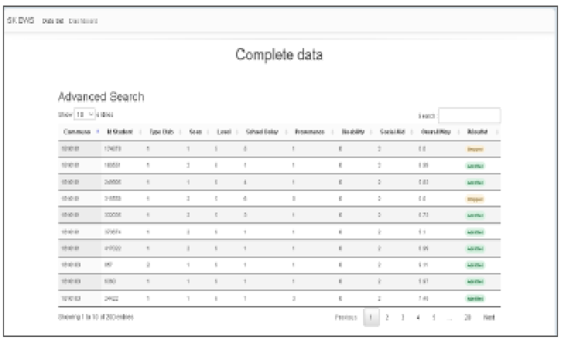


FIGURE 8. Data Set table with notification of prediction results.

For instance, a student who has benefitted from pre-school education gets a head start when it comes to understanding and completing primary school subjects, and a psychological edge because of the familiarity with the school and classroom environment. Because more academic failures accrue and the notion of quitting school to pursue other choices is firmly present the higher a child climbs, level also plays a role in the dropout equation.

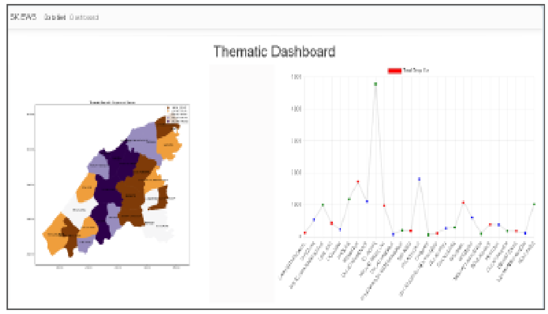


FIGURE 9. Visualization of the most affected areas by the drop-out phenomenon and their classification.

There are a few other characteristics that may be thought of as having a less substantial impact on the label class prediction. A student who receives some kind of social assistance, such as financial help, may still be regarded to be from a disadvantaged and vulnerable background, although the model did not place a high value on this factor. Since "housing environment" inevitably relates to social status and the availability of material means, the area is the other aspect of the idea of social support. The characteristic Type of school is likewise ranked poorly by the model. The sociology of education discusses the school's impact and must take into account the fact that students attending "unlucky schools" (those located in low-income or otherwise disadvantaged areas) are more prone to academic failure and, ultimately, dropout. Students with special needs have nearly the same chance of continuing their school career or dropping out as regular students, possibly due to inclusive pedagogy and classrooms that cater to these stu-dents with special needs, which is the last attribute in the importance ranking that is Disability.

DISCUSSION

A number of programs and efforts have sought to compile all the elements that contribute to student dropout, and this is an extremely essential point to make. Data collection, processing, and interpretation pose significant challenges in the educational sector, as they do in other scientific domains. To include the many scientific aspects of education, the phrase "educational sciences" has only just been developed.
In 2014, Lee et al. [8] captured a snapshot of the South Korean National Education Information System (NEIS) database. High school dropouts were predicted using 15 factors spanning several behavioral, familial, academic, and other domains. This was particularly effective in avoiding the problems associated with class-imbalanced datasets by using the boosted decision tree (BDT) method with minority oversampling techniques (SMOTE), which led to very significant findings (a 99% prediction accuracy for the Target attribute). The model's use of binary classification does improve the discriminatory power of the findings, but it also obscures, from a different perspective, all the subtleties that might still identify the students' circumstances. The dropout

prediction battery, on the other hand, paid little attention to students' demographic and personal characteristics and instead concentrated on so-called "external" and "environmental" variables.

TABLE 5. Benchmarking of our project results.

	Analysis factors					
	Demographic	Socio economic	Macro economic	Academic	Environmental	Applied Method
Our data	x	x		x	x	KNN (99%)
PIP data	x	x	x	x		RF (90%)

A Gradual At-Risk (GAR) model has been put into place by Bañeres et al. [26] using data collected online from the Universitat Oberta de Catalunya (UOC) datamart. This project's overarching goal is to identify college freshmen who have the highest probability of failing a certain course and, ultimately, of dropping out of school altogether. Students' actions on UOC operational systems and aggregated course assessment data over lengthy periods of time are the data used. For every class, the GAR model our research developed outperformed the competition using the K-Nearest Neighbors technique, which achieved an accuracy rate of 95% or above. The majority of student modeling initiatives in the realm of online education have relied on past data collected from students' engagement with a certain course or curriculum. Unfortunately, there is yet not enough data in this database to do a comprehensive analysis of what causes a student to withdraw from a class. Similar to the UOC, which targets students in higher education, other educational cycles and levels go unnoticed and uninterested due to the difficulty of reaching this kind of student, particularly in light of the globalization of primary education. In a similar vein, Realinho et al. [37] created a unified dataset by merging data from several systems at Portugal's Polytechnic Institute of Porto (PIP). Academic trajectory, demography, macroeconomics, socioeconomic determinants, and grade-level achievement were among the many elements they sought to add to the database. Determine the most essential factors weighing on the prediction model to raise the accuracy of the system and successfully anticipate the dropout phenomenon for all students enrolled in the institute's academic disciplines. The database here is a crude reflection of the concept we're attempting to articulate with this work. Our database has been built with the goal of include all potential explanatory aspects for the dropout phenomenon. Realinho et al.'s [37] academic database allowed researchers to investigate academic dropout rationale in great detail, and the data set was improved by a number of changes. Regarding the analytical components used for prediction and the logic for processing the issue, there were several parallels to our test. Table 5 shows the similarity between the two databases used in the quest to model the dropout phenomenon.

CONCLUSIONANDFUTUREWORK

Education is a very nuanced matter of strategic national security since a country's development is directly related to the literacy and training levels of its people. Those in charge of education are therefore faced with a dilemma when considering the issue of school dropouts; some see it as a waste of resources and an investment that does not benefit future generations. Given the complexity of the issue at hand, we shall narrow our attention on primary cycle data as they relate to early school wastage. This problem persists in contexts with limited resources, despite several efforts to resolve it. Our EWS project's main objective is to identify the groups and individuals who would be most affected by the issue of school dropout and to shed light on its many causes. These indicators should be taken into account by educational planners so that interventions and policies may be targeted towards the areas where the phenomenon is most prevalent.

No matter how well the system is operating, there is always opportunity for improvement. The expansion of the Data Set, both in terms of the amount of data and, more crucially, the addition of additional characteristics that provide increased precision of the model's forecast. Due to the data's dispersion across multiple operating systems and formats, it was tough to incorporate all essential indicators in the Data Set we prepared for this project, although we made every attempt to do so. Another part of development is improving our client app by adding new admin and data manipulation interfaces that are dynamic and interactive. Finally, we may look at creating a recommendation system that suggests appropriate solutions for future situations based on the EWS's findings.

REFERENCES

[1] Y. Han and S. Liu, ''Construction and research of big data platform forparty building in colleges and universities,'' in *Proc. 2nd Int. Conf. Internet,Educ. Inf. Technol.*, 2022, pp. 431–436, doi: 10.2991/978-94-6463-058-9_70.

[2] A. N. de Vasconcelos, L. A. Freires, G. D. L. Loureto, G. Fortes,J. C. A. da Costa, L. F. F. Torres, I. I. Bittencourt, T. D. Cordeiro, and S.Isotani, ''Advancing school dropout early warning systems: The IAFREErelational model for identifying at-risk students,'' *Frontiers Psychol.*,vol. 14, Jul. 2023, Art. no. 1189283, doi: 10.3389/fpsyg.2023.1189283.

[3] B. M. Mcmahon and S. F. Sembiante, ''Re-envisioning the purpose ofearly warning systems: Shifting the mindset from student identificationto meaningful prediction and intervention,'' *Rev. Educ.*, vol. 8, no. 1,pp. 266–301, Feb. 2020, doi: 10.1002/rev3.3183.

[4] Z. Alharbi, J. Cornford, L. Dolder, and B. De La Iglesia, ''Usingdata mining techniques to predict students

at risk of poor performance,''in *Proc. SAI Comput. Conf. (SAI)*, Jul. 2016, pp. 523–531, doi:10.1109/SAI.2016.7556030.
[5] M. S. Ahmad, A. H. Asad, and A. Mohammed, ''A machine learningbased approach for student performance evaluation in educational datamining,'' in *Proc. Int. Mobile, Intell., Ubiquitous Comput. Conf. (MIUCC)*,May 2021, pp. 187–192, doi: 10.1109/MIUCC52538.2021.9447602.
[6] Y.-H. Hu, C.-L. Lo, and S.-P. Shih, ''Developing early warning systemsto predict students' online learning performance,'' *Comput. Hum. Behav.*,vol. 36, pp. 469–478, Jul. 2014, doi: 10.1016/j.chb.2014.04.002.
[7] Y.-C. Chang, W.-Y. Kao, C.-P. Chu, and C.-H. Chiu, ''A learning styleclassification mechanism for e-learning,'' *Comput. Educ.*, vol. 53, no. 2,pp. 273–285, Sep. 2009.
[8] S. Lee and J. Y. Chung, ''The machine learning-based dropout earlywarning system for improving the performance of dropout prediction,''*Appl. Sci.*, vol. 9, no. 15, p. 3093, Jul. 2019, doi: 10.3390/app9153093.
[9] S. Bansal and N. Baliyan, ''A study of recent recommender system techniques,''*Int. J. Knowl. Syst. Sci.*, vol. 10, no. 2, pp. 13–41, Apr. 2019, doi:10.4018/ijkss.2019040102.
[10] J. Mostow, J. Beck, H. Cen, A. Cuneo, E. Gouvea, and C. Heiner, ''An educationaldata mining tool to browse tutor-student interactions: Time willtell,'' in *Proc. Workshop Educ. Data Mining, National Conf. Artif. Intell.*,2005, pp. 15–22.
[11] M. A. M. Iver and D. J. M. Iver, ''Beyond the indicators: An integratedschool-level approach to dropout prevention,'' George WashingtonUniv. Center Equity Excellence Educ., Arlington, VA, USA, Tech. Rep.ED543512, 2009.
[12] U. B. Qushem, S. S. Oyelere, G. Akcapinar, R. Kaliisa, and M. J. Laakso,''Unleashing the power of predictive analytics to identify at-risk students incomputer science,'' *Technol., Knowl. Learn.*, vol. 28, pp. 1–16, Jul. 2023,doi: 10.1007/s10758-023-09674-6.
[13] W. M. Ei Leen, N. A. Jalil, N. M. Salleh, and I. Idris, ''Dropout earlywarning system (DEWS) in Malaysia's primary and secondary education:A conceptual paper,'' in *Proc. Int. Conf. Inf. Syst. Intell. Appl.* (LectureNotes in Networks and Systems), vol. 550, 2023, pp. 427–434, doi: 10.1007/978-3-031-16865-9_33.
[14] M. Yagci, ''Educational data mining: Prediction of students' academicperformance using machine learning algorithms,'' *Smart Learn. Environments*,vol. 9, no. 1, Dec. 2022, Art. no. 11, doi: 10.1186/s40561-022-00192-z.
[15] X. Sun, Y. Fu, W. Zheng, Y. Huang, and Y. Li, ''Big educationaldata analytics, prediction and recommendation: A survey,'' *J. Circuits,Syst. Comput.*, vol. 31, no. 9, Jun. 2022, Art. no. 2230007, doi:10.1142/S0218126622300070.
[16] T. A. Kustitskaya, A. A. Kytmanov, and M. V. Noskov, ''Early studentat-risk detection by current learning performance and learning behaviorindicators,'' *Cybern. Inf. Technol.*, vol. 22, no. 1, pp. 117–133, Mar. 2022,doi: 10.2478/cait-2022-0008.
[17] Y. Wu, ''Design framework of early warning mechanism of studentachievement with machine learning,'' in *Proc. IEEE 4th EurasiaConf. IoT, Commun. Eng. (ECICE)*, Oct. 2022, pp. 239–241, doi:10.1109/ECICE55674.2022.10042897.
[18] Y. Qu, Z. Sun, and L. Liu, ''Research on the academic early warning modelof distance education based on student behavior data in the context ofCOVID-19,'' *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 10, pp. 633–642,2022, doi: 10.14569/IJACSA.2022.0131074.
[19] S. Yi, ''The development of early warning system for college Englishacademic performance based by big data computation,'' in *Proc. Int.Conf. Comput., Big-Data Eng. (ICCBE)*, May 2022, pp. 226–229, doi:10.1109/ICCBE56101.2022.9888191.
[20] A. B. Soussia, A. Roussanaly, and A. Boyer, ''Toward an early risk alertin a distance learning context,'' in *Proc. Int. Conf. Adv. Learn. Technol.(ICALT)*, Jul. 2022, pp. 206–208, doi: 10.1109/ICALT55010.2022.00067.