Journal of Cardiovascular Disease Research
ISSN: 0975-3583, 0976-2833 VOL15, ISSUE12, 2024

DEVELOPMENT OF AN EARLY WARNING SYSTEM TO SUPPORT EDUCATIONAL PLANNING PROCESS BY IDENTIFYING AT-RISK STUDENTS

MANASU MADHAVI[1], CH. RAMYA[2], D. HIMA VARSHA[3], B. LAXMI[4]
ASSISTANT PROFESSOR 1, UG STUDENT 2,3,4
DEPARTMENT OF IT, MALLA REDDY ENGINEERIGNG COLLEGE FOR WOMEN (UGC-AUTONOMOUS), MAISAMMGUDA, HYDERABAD, TELANGANA-500100

ABSTRACT

The development of data analysis techniques and intelligent systems has had a considerable impact on education, and has seen the emergence of the field of educational data mining (EDM). The Early Warning System (EWS) has been of great use in predicting at-risk students or analyzing learners' performance. Our project concerns the development of an early warning system that takes into account a number of socio-cultural, structural and educational factors that have a direct impact on a student's decision to drop out of school. We have worked on an original database dedicated to this issue, which reflects our approach of seeking exhaustiveness and precision in the choice of dropout indicators. The model we built performed very well, particularly with the K-Nearest Neighbor (KNN) algorithm, with an accuracy rate of over 99.5% for the training set and over 99.3% for the test set. The results are visualized using a Django application we developed for this purpose, and we show how this can be useful for educational planning.

INTRODUCTION

The evolutionary path of IT practice has taken a new form with the advent of intelligent systems, especially predictive and recommendation systems. And with the explosion of data and the entry into the era of Big Data [1], these systems have found more opportunities to flourish and achieve the most remarkable results. Early warning systems (EWS) are one of the most famous types of intelligent systems, and have benefited from the considerable leap forward in computing methods and technologies used, as well as the development of hardware infrastructures. The EWS is a predictive system that aims to support decision-making by giving a proactive view of the future situation by analyzed data. EWS are used in almost all fields, and their role lies in detecting anomalies in real systems and warning decision-makers of the seriousness of situations, so that they can anticipate their intervention to remedy the problems posed, or at least limit the negative effects and consequences. An EWS can be defined by a number of active keywords [2], [3]: Collect, Analyze, Detect, Prevent, Alert, Notify. Each word indicates one of the key stages of an EWS, hence its action model, which consists of a set of layers or steps. The first step involves the continuous monitoring of relevant indicators and the collection of data in real or nearreal time. The next step involves the analysis and processing of the data collected. This process involves examining the data to identify early indicators, patterns or deviations from the norm. Various techniques such as advanced algorithms, statistical models or artificial intelligence can be used to identify potential anomalies during this analysis. The third step is 'Alert and Notify', once the system detects an irregularity or potential hazard, it quickly triggers an alert to inform the parties concerned. Then, in the fourth step 'Risk assessment',professionals and supervisors evaluate the reliability and severity of the warning. They examine existing data and information to understand the characteristics of the risk, its potential consequences and possible actions to minimize itsimpact. The communication and dissemination process plays an essential role in risk management and response. Once a risk has been assessed and verified, it is crucial to share the relevant information with all parties concerned, including stakeholders, decision-makers and the general public. The final layer of the process involves response and action. And finally, once the early warning system provides the necessary information, appropriate measures are taken to reduce risks, prevent crises or minimize adverse effects. All these steps form an iterative process, as we are always aiming for perfection of the system, given that EWSs are used in highly critical areas and that the effect of their Outputs can avert disasters in some cases, whether in the near future or in the long term. That's why we're constantly striving for perfection.

LITERATURE REVIEW

Use of Utility Based Interestingness Measures to Predict the Academic Performance of Technology Learners in Sri Lanka

- K. Kasthuriarachchi, S. Liyanage
- Published in International Conference on… 1 August 2018

Knowledge extracted from educational data can be used by the educators to obtain insights about how the quality of teaching and learning must be improved, how the factors a $\square$ ect the performance of the students and how qualified students can be trained for the industry requirements. This research focuses on classifying a knowledge based system using a set of rules. The main purpose of the study is to analyse the most influencing attributes of the students for their module performance in tertiary education in Sri Lanka. The study has gathered data about students in a reputed degree awarding institute in Sri Lanka and used three different data mining algorithms to predict the influential factors and they have been evaluated for interestingness using objective oriented utility based method. The findings of this study will positively a $\square$ ect the future decisions about the progress of the students' performance, quality of the education process and the future of the education provider.

An Efficient Approach of Feature Selection and Metrics for Analyzing the Risk of the Students Using Machine Learning

- V. B. Gladshiya, K. Sharmila
- Published in International Conference on… 8 October 2021

An eminent fortune in this world today is Data. A million bytes of data can be originated every day in every field of applications such as education, medical, business, government, organizations etc. In educational field the performance of the students is the midway so that it is essential to identify the risk of the students using predictive analytics. If the student risk is identified their performance can be predicted using machine learning algorithms. Machine Learning algorithms are the tools that process the data in an efficient aspect and play a multimodal feature in the field of Data Science, Artificial Intelligence, Predictive analytics etc. For data analytics the machine learning algorithms could not find the text, image or video. Hence it is essential to preprocess the data set which can be then used for analytics for identifying future inferences using Machine learning algorithms. A data set is a collection of samples or objects which can be characterized by the features called variables or attributes. The features are the key elements of the data sets through which the predictions can be obtained by selecting the specified features for correlated prediction. This paper expounds the working methods of data preprocessing and feature selection for predicting the student's performance and also compares the metrics with its threshold value which would be used for future research work.

Towards Finding a Minimal Set of Features for Predicting Students' Performance Using Educational Data Mining

- S. Sengupta
- Published in International Journal of… 8 June 2023

An early prediction of students' academic performance helps to identify at-risk students and enables management to take corrective actions to prevent them from going astray. Most of the research works in this field have used supervised machine learning approaches to their crafted datasets having numerous attributes or features. Since these datasets are not publicly available, it is hard to understand and compare the significance of the chosen features and the efficacy of the different machine learning models employed in the classification task. In this work, we analyzed 27 research papers published in the last ten tears (2011-2021) that used machine learning models for predicting students' performance. We identify the most frequently used features in the private datasets, their interrelationships, and abstraction levels. We also explored three popular public datasets and performed statistical analysis like the Chi-square test and Person's correlation on its features. A minimal set of essential features is prepared by fusing the frequent features and the statistically significant features. We propose an algorithm for selecting a minimal set of features from any dataset with a given set of features. We compared the performance of different machine learning models on the three public datasets in two experimental setups-one with the complete feature set and the other with a minimal set of features. Compared to using the complete feature set, it is observed that most supervised models perform nearly identically and, in some cases, even better with the reduced feature set. The proposed method is capable of identifying the most essential feature set from any new dataset for predicting students' performance.

EXISTING SYSTEM Our contribution in this paper intersects with several initiatives and projects that have been working on the implementation of EWS in the education system. And EWS is a field that has prompted a great deal of production in recent years. Researchers in the field of education, whose aim is to develop teaching practice and minimize the risk of student attrition, have made a number of efforts on several fronts [11]. The various EWSs have acted on various factors, depending on the logic of analysis, the field of practice or the purpose of the system. These factors could be classified as follows:

• School factors: Represent the different variants that affect students' school performance. On the one hand, we start with the pupil's educational background, i.e.: everything to do with where he or she comes from, did he or she receive pre-school education or not? did he or she come from formal, non-formal or original education or something else? On the other hand, his or her learning experience throughout the years of schooling prior to the time of the study (grades and averages for courses and exams, as well as end-of-year averages). These factors represent a very important part of a student's heritage, especially if they are analyzed over time, in a cumulative way, and if possible within the frame of a cohort. In most cases, this time-series study is not possible, mainly due to a lack of data, which leads researchers to make projections on the basis of a school season. But in any case, school factors alone do not allow for an in-depth study of educational phenomena, perhaps only a statistical analysis of school performance.

•Human factors [5]: All the elements that define the learner as a human being, by his gender, age, ethnicity or origin. At this level, we also talk about students' behavioural traits and habits, especially those that affect their learning and reduce their academic performance: absenteeism, procrastination, diligence. . .Body and mental aspects are also taken into consideration, since body health (disabled or not) and mental health (sick or healthy) can significantly affect a learner's performance. These factors are of major importance in the study of the learner's being and personality. Behavioural data can be collected through empirical studies or only through individual analysis of learners. Nevertheless, the lack of medical data is very much in evidence in the systems of underdeveloped countries, resulting in the marginalization of a very important effect that may explain the difficulty or deterioration of the learner's performance or retention. But added to this, the human being does not live in a personal context alone but in a society and environment, which likewise makes human indicators insufficient in the process of predicting educational phenomena.

• Environmental factors [15],: The effect of the environment is undoubtedly a very significant one on the individual, which is why we speak in the education's sociology of a fortunate school belonging to a resource-rich environment, and an underprivileged school located in an underprivileged environment. The type of school is also categorized at this level, since a private school is often better in terms of academic results than its public counterpart, mainly due to disparities in resources. Another essential dimension of these factors is the family's internal situation, or as it is known in the general population census: household status. The latter highlights indicators of the family's economic capacity and social status, as well as the parents' level of education and the culture that reigns within the family and the aspirations of its members (parents, brothers and sisters). Finally, the school's internal atmosphere is also important in this context:

ISSN: 0975-3583, 0976-2833 VOL15, ISSUE12, 2024

the school's management model, facilities, the state of the teacher/learner relationship, internal peace and non-violence are maintained. These factors highlight the way in which learners are impacted by their school environment, as well as the actors operating in the school. On the other hand, we need to be careful about the way in which the environment is apprehended, as well as the type of variables used, for example, the degree of family impact, in relation to the school or the street. Also, perhaps there are other environmental factors that we couldn't capture but that are very important too, the effect of media and social networks, or that of older students in higher grades on new entrants, and others.

• Exceptional factors [6]: These include all exceptional contingencies and natural crisis situations such as natural disasters, wars and human conflicts that destabilize the educational process. One very exceptional situation we have witnessed over the last 3 years is the health pandemic caused by the Corona virus. This pandemic had an extraordinary effect on the whole learning mechanism, with schools closing their doors all over the world, and other learning methods taking over from face to- face learning, such as distance learning, where e-learning and hybrid methods flourished. These circumstances usually have a radical and abrupt effect, and can only be studied in exceptional vision. But they are not taken into account in the development of a general model aimed at sustainability.

DISADVANTAGES:

• The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets to detect and Identifying At-Risk Students.

• Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer.

• Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions..

PROPOSED SYSTEM

We used mainly machine learning analysis methods specialized in classification. Hence the use of the best-known and most efficient for this aspect of analysis: SVM, Random Forest, SGD and KNN.We have tried to calibrate the internal parameters of each algorithm in order to use the best possible version.

1) SVM

Supervised machine learning encompasses a range of algorithms, one of which is the support vector machine (SVM). This algorithm is specifically designed to handle classification and regression tasks. Its effectiveness is particularly pronounced when dealing with complex datasets containing many features. The main objective of SVM is to locate a hyperplane, which serves as a decision boundary, that optimally separates distinct classes of data points.

2) RANDOM FOREST

Random Forest is a popular ensemble learning algorithm used for classification and regression tasks. It aims to improve the accuracy and robustness of predictions by combining the output of several individual models (usually decision trees). The algorithm takes its name from the idea of creating a ''Forest'' of decision trees, each trained on a random subset of the data. The idea behind a Random Forest is to reduce over fitting and improve generalization by leveraging the diversity of multiple decision trees.

3) SGD

It stands for Stochastic Gradient Descent, which is a popular optimization algorithm used in machine learning and deep learning for training models, particularly when large datasets are involved. SGD is a variant of the more general gradient descent optimization algorithm, but introduces randomness and speed improvements that make it suitable for large-scale data processing.

4) KNN

K-Nearest Neighbours (KNN) is a simple and intuitive machine learning algorithm for classification and regression tasks. It is instance-based learning that makes predictions based on the majority class or average of the k nearest data points in the feature space.KNN algorithm is calculated using various distance metrics. The two most commonly used distance metrics are Euclidean distance and Manhattan distance. These metrics provide a way to measure the similarity or dissimilarity between two data points in the feature space

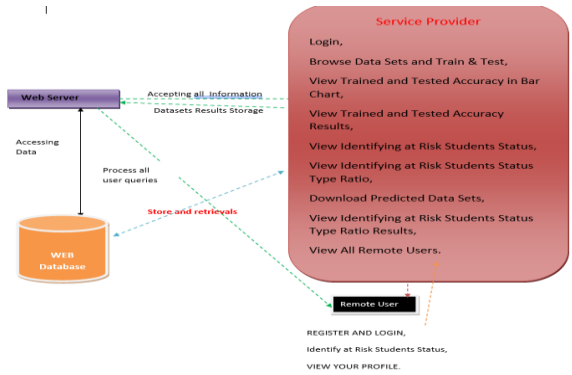
ADVANTAGES:

The EWS will also support a process of data and model evaluation, and we will use a set of the most effective metrics to be confident of the reusability of our model. The danger of this maneuver is that any error in the prediction will automatically lead us to make a mistake about the future of one or more students, resulting in the wrong intervention and the wrong decision for the situation under study.

IMPLEMENTATION

SYSTEM ARCHITECTURE

ISSN: 0975-3583, 0976-2833 VOL15, ISSUE12, 2024



MODULES

- **SERVICE PROVIDER**

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Identifying at Risk Students Status, View Identifying at Risk Students Status Type Ratio, Download Predicted Data Sets, View Identifying at Risk Students Status Type Ratio Results, View All Remote Users.

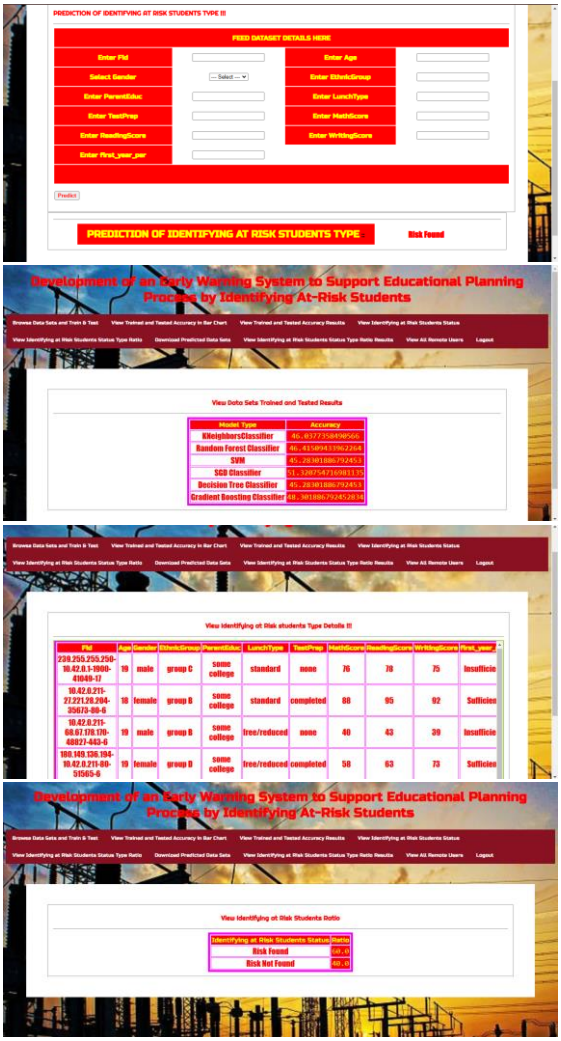
- **VIEW AND AUTHORIZE USERS**

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

- **REMOTE USER**

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, Identify at Risk Students Status, VIEW YOUR PROFILE.

RESULT



CONCLUSION

Education is a highly sensitive area of strategic national security, since the development of any country has a direct bearing on the level of literacy and training of its citizens. That's whythe phenomenon of school drop-outs represents a dilemma for those in charge of education, and is seen as a drain on financial resources and a waste of effort with no return on investment for those generations who don't continue their learning. This situation is seen as even more delicate when itcomes to early school wastage, as in the primary cycle, which is why we have chosen to focus on data from this cycle. Despite all the initiatives aimed at tackling this phenomenon, it still persists, especially in contexts where there is a lack of resources. Our EWS project aims to provide a panoramic view of the various factors that can cause schooldropout, as well as a proactive vision of the areas most affected by this phenomenon, or the individuals likely to be victims of dropout. These indicators will need to be taken on board by educational planners in order to target areas where the phenomenon is most prevalent with appropriate policiesand interventions. Although the system is performing well, there is still considerable scope for development. A first dimension is the expansion of the Data Set, both in terms of the data quantity, but above all the addition of other attributes that give greater precision to the predictive model. For the Data Set we developed for this project, we did our best to gather as many indicators as

possible, but this wasn't easy given thedifficulty of accessing the information and its dispersion over several operational systems with diversified forms. Another dimension of evolution is the upgrading of our client application by adding further interfaces for administration and data manipulation in dynamic and interactive ways. Finally, we may work on a recommendation system that takes into account the outputs of the EWS in order to propose appropriate interventions in predicted situations.

REFERENCES

[1] Y. Han and S. Liu, ''Construction and research of big data platform for party 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 IAFREE relational 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 of early warning systems: Shifting the mindset from student identification to 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, ''Using data 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 learning based approach for student performance evaluation in educational data mining,'' 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 systems to 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 style classification 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 early warning 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 educational data mining tool to browse tutor-student interactions: Time will tell,'' 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 integrated school-level approach to dropout prevention,'' George Washington Univ. 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 in computer 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 early warning system (DEWS) in Malaysia's primary and secondary education A conceptual paper,'' in Proc. Int. Conf. Inf. Syst. Intell. Appl. (Lecture Notes 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' academic performance 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 educational data analytics, prediction and recommendation: A survey,'' J. Circuits, Syst. Comput., vol. 31, no. 9, Jun. 2022, Art. no. 2230007, doi: 10.1142/S0218126622300070.