AN INTELLIGENT EDUCATION SYSTEM FOR ONLINE PLATFORM USING MACHINE LEARNING APPROACH

Hamad Abed Farhan
hamdabd946@gmail.com
The Republic of Iraq Sunni Endowment Office/
Department of Religious Education and Islamic Studies
Research Department

ABSTRACT

The utilization of Educational Data Mining (EDM) and learning analytics has become paramount in enhancing the quality of result classification in student performance analysis. This paper proposes a comprehensive framework aimed at improving student outcomes through three distinct phases: Cluster-based Student Record Classification (CSRC), Multi-Tier Student Performance Evaluation Model (MTSPEM), and Behaviour-based Student Classification System (SCS-B). The framework begins with data collection from the student database, focusing on academic and learning-related parameters. Subsequently, K-Means Clustering is employed for efficient data organization and classification. The selected attributes are then transformed and processed for accurate student classification, considering factors such as grade points, attendance, and educational background. The J48 tree-based classification model is utilized for precise categorization of student performance. Additionally, K-Means Clustering is applied to further refine data clustering and prediction accuracy. The proposed model is implemented using WEKA, an open-source data analysis tool, and evaluated against existing classification techniques such as Support Vector Machine (SVM) and Multi-layer Perceptron (MLP). The results demonstrate superior performance of the proposed framework, with classification accuracy reaching up to 98%. This research provides valuable insights for educational institutions to tailor interventions and support mechanisms for students, ultimately leading to improved academic outcomes and student success.

Keys Words: Machine Learning, Data Science, Learning Analytics, Educational Informatics, Artificial Intelligent in Education.

نظام تعليمي ذكي لمنصة عبر الإنترنت باستخدام أسلوب التعلم

حمد عبد فرحان

جمهورية العراق ديوان الوقف السني/ قسم التربية الدينية والدراسات الإسلامية

قسم البحوث

الملخص

أصبح استخدام استخراج البيانات التعليمية (EDM) وتحليلات التعلم أمرًا بالغ الأهمية في تعزيز جودة نتائج التصنيف في تحليل أداء الطلاب. تقترح هذه الورقة إطارًا شاملاً يهدف إلى تحسين نتائج الطلاب من خلال ثلاث مراحل متميزة: تصنيف سجلات الطلاب على أساس المجموعة (CSRC)، ونموذج تقييم أداء الطلاب متعدد المستويات (MTSPEM)، ونظام تصنيف الطلاب على أساس السلوك (SCS-B). يبدأ الإطار بجمع البيانات من قاعدة بيانات الطلاب، مع التركيز على المعلمات الأكاديمية والمتعلقة بالتعلم. وبعد ذلك، يتم استخدام K-Means Clustering لتنظيم البيانات وتصنيفها بكفاءة. يتم بعد ذلك تحويل السمات المحددة ومعالجتها للحصول على تصنيف دقيق للطلاب، مع الأخذ في الاعتبار عوامل مثل درجات التقدير والحضور والخلفية التعليمية. يتم استخدام نموذج التصنيف القائم على الشجرة J48 للتصنيف الدقيق لأداء الطلاب. بالإضافة إلى ذلك، يتم تطبيق K-Means Clustering لزيادة تحسين تجميع البيانات ودقة التنبؤ. يتم تنفيذ النموذج المقترح باستخدام WEKA، وهي أداة تحليل بيانات مفتوحة المصدر، ويتم تقييمه وفقًا لتقنيات التصنيف الحالية مثل Support Vector Machine (SVM) وMulti-layer Perceptron (MLP). أظهرت النتائج أداءً متفوقًا للإطار المقترح، حيث وصلت دقة التصنيف إلى ٩٨%. يوفر هذا البحث رؤى قيمة للمؤسسات التعليمية لتصميم التدخلات وآليات الدعم للطلاب، مما يؤدي في النهاية إلى تحسين النتائج الأكاديمية ونجاح الطلاب.

الكلمات المفتاحية: التعلم الآلي، علم البيانات، تحليلات التعلم، المعلوماتية التعليمية، الذكاء الاصطناعي في التعليم.

INTRODUCTION

In recent decade, there is an increasing interest and demand for models detecting the influencing factors of student performances in education, specifically incorporating data mining models. The data mining in educational domain and research is termed as Educational Data Mining (EDM) in [1]. The main motive of Educational Data Mining is to help for identifying the poor performing students earlier and improving their learning skills, which makes the institution to provide higher educational standards. Moreover, Educational Data Mining is the fast growing research domain because of its capability to obtain information from large amount of student data worked by [2]. In recent times, data mining and student information systems are effectively integrated for evaluating the student performances accurately in [3]. Additionally, the Educational Institutions use this for digitalizing the student data and transactions in [4]. The huge amount of student information, presented in Student Information System, includes the following factors, i. Course Data ii. Tutor Data iii. Student Personal Data iv. Student Demographics v. Student's Grade vi. Attendance Data

In EDM, the applications and services of data mining operations, machine learning and analytics to the data are observed from educational data bases from universities. At the top level, the domain searches to frame and enhances the techniques for data exploration that applies in hierarchical operations in multiple levels. Based on that, the EDM techniques has provided two methods of learning analytics by the research people in learning psychology of students. EDM is concerned with application and development of mining techniques for pattern recognition from large educational dataset and for better student and environmental analysis done by [5]. Educational Data Mining incorporates various techniques of data mining and data analytics carried out for processing. The frequently used prediction methods are classification, regression and latent feature derivation methods. Unsupervised learning methods such as clustering, factor evaluations and network computations are used for determining the efficient structure in several domains of educational data.

STUDENT WELL-BEING

Student well-being encompasses the overall positive state of individuals in handling the challenges of student life and learning. Unlike mental health issues, which affect specific individuals and manifest as conditions like depression, anxiety, or isolation, well-being addresses the broader population's ability to cope effectively [6]. Previous studies have indicated a correlation between poor well-being, mental health issues, and low academic performance [7]. Additionally, interventions aimed at enhancing student well-being have demonstrated a positive impact on academic outcomes, highlighting the interconnectedness between mental health and academic achievement [8]. A research endeavor examining the mental health status of secondary school students in Canada [9] delved into the correlation between symptoms of mental illness (e.g., depression and anxiety) and mental well-being with self-reported grades and educational behaviors. Utilizing fundamental factual examinations to check heading, size, and importance, the concentrate reliably uncovered that decreased sadness levels and increased prosperity were connected with worked on scholastic execution and instructive ways of behaving. Furthermore, other research findings underscore the significant variability in students' well-being across different schools, emphasizing that institutions implementing well-being promotion practices tend to exhibit higher overall well-being levels [10]. Similarly, studies in the realm of higher education echo these findings. A UK-based examination [11] investigated the perplexing associations among wellbeing mindfulness, wellbeing conduct, emotional wellbeing status, fulfillment with instructive experience, and different proportions of scholastic accomplishment. Integrating components connected with prosperity like satisfactory pay, rest quality, and actual wellbeing, the outcomes revealed a corresponding connection between wellbeing, wellbeing ways of behaving, and instructive fulfillment. One more concentrate in Romania [12] dove into the nexus between scholarly execution, understudy commitment, and burnout, revealing insight into the complex elements affecting understudy prosperity and instructive results. This study examined the potential causal pathways between grades and student well-being, considering both scenarios: grades as an outcome of student well-being and student well-being as an outcome of academic performance (grades). A noteworthy revelation from this investigation is the identification of high academic grades as a precursor to increased student engagement and reduced student burnout. Conversely, the study observed that burnout did not exert a significant impact on subsequent academic performance, highlighting the complex interplay between student well-being and educational outcomes.

ETHICAL LEARNING ANALYTICS

Educational platforms and digital tools provide a wealth of data that opens up new avenues for analyzing students' learning approaches. This data not only offers fresh perspectives on how teaching and learning can be improved but also enhances the likelihood of learner success. However, while there is significant emphasis on leveraging technology and data processing techniques in Learning Analytics (LA), ethical considerations often lag behind. The majority of research in LA primarily focuses on developing methods and techniques for analyzing intricate data sets, with insufficient attention given to the ethical implications. A literature review conducted in 2018 on learning analytics revealed that only 18% of the articles examined mentioned or considered ethics and privacy concerns in their research endeavors to varying extents.

The partners engaged with instructive innovation, including instructive organizations, instructors, understudies, and EdTech organizations, frequently harbor unique qualities and interests. Instructive organizations bear lawful and moral obligations to sustain understudies' prosperity and work with their scholarly achievement. Learning Analytics (LA) can serve as a valuable tool for schools and educators, aiding classroom practices. However, concerns have been raised that initiatives like well-being campaigns may strain existing resources. Conversely, EdTech companies, often constrained by limited funding, are driven to swiftly amass extensive data to enhance their products and algorithms. As technological advancements rapidly unfold and substantial financial interests are at play, developers and researchers of Learning Analytics (LA) technology often prioritize speedy releases to outpace competitors, relegating ethical and societal considerations to secondary importance. Moreover, the power dynamics between students and educational institutions introduce asymmetries that cannot be overlooked. Educational institutions possess access to an increasing volume and diversity of student data, thereby bearing a significant responsibility to ensure the ethical collection, utilization, and storage of such data. The introduction of the General Data Protection Regulation (GDPR) by the European Union in 2018 underscores the importance of regulating the processing of user data, extending its implications to organizations, including educational institutions.

RELATED WORKS

A comprehensive examination of educational data mining and learning analytics conducted by [19] surveyed current research in higher education, revealing prevalent methods of data analysis in learning analytics, notably prediction, clustering, and relationship mining [20]. Among the primary tasks addressed by learning analytics research, forecasting student performance stands out, with regression and classification emerging as the predominant methods [21]. Notably, [21] underscores the scarcity of data available for primary and secondary education compared to tertiary levels, with the latter referring to university and college education. The review indicates a significant skew in research focus, with 86.6% of published papers centered on tertiary education, a mere 7.3% on secondary education, and none dedicated to primary education. An underlying factor contributing to this trend may stem from the greater digitalization of data in university education compared to primary and secondary education, making data more accessible for analysis. However, this underscores the significance of expanding learning analytics research to encompass primary and secondary education levels. Predicting academic performance emerges as a predominant focus within Learning Analytics, denoting the scores achieved by students in assessments conducted at the conclusion of a learning period. The term "Early Warning System" commonly denotes a framework leveraging student data variables to forecast their academic performance or assess the risk of dropout [22] [23].

During a longitudinal investigation spanning from 2009 to 2012 in the United States [24], Machine Learning techniques were employed to identify students at risk of dropping out from upper secondary school. Notably, the study highlighted key variables essential for the predictive model, emphasizing the significance of student GPA, age, math test scores, expulsion records, and attendance [25]. In a systematic review addressing the Early Prediction of Student Learning Performance [26], diverse Machine Learning methodologies and predictor variables were examined across various studies. The review underscored the variability in prediction accuracy across these studies, underscoring the crucial role of the number and types of variables integrated into the prediction models. Across different educational settings, the predictors and attributes utilized for predictive modeling exhibited variability. Typically, these factors encompassed student demographics, activities, and interactions within an eLearning framework. Notably, an important observation is the absence of student well-being as a variable in the predictive models examined in the aforementioned studies, indicating the need for additional research in this domain.

PROPOSED SYSTEM

The Educational Data Mining (EDM) and learning analytics are effectively used for enhancing the quality of result classification in student performance analysis. The educational institutions are involved in things to reduce the poor results of students. With that concern, many techniques are developed for evaluating the student performances for making the respective faculties to mediate to improve the overall results. For developing Accurate Student Classification Model, this work comprises of three phases of work, as follows:

i. Cluster based Student Record Classification (CSRC)

ii. Multi-Tier Student Performance Evaluation Model (MTSPEM)

iii. Behaviour based Student Classification System (SCS-B)

The working procedure of the aforementioned models and computations are presented in the following sections. The overall functions involved in the three phases of work are presented in the following Figure 1. In EDM, the huge amount of student data is collected through various survey as well as from open source data repository. For efficient data organization, analysis and classification, K-Means Clustering techniques are used with respect to the obtained academic data.

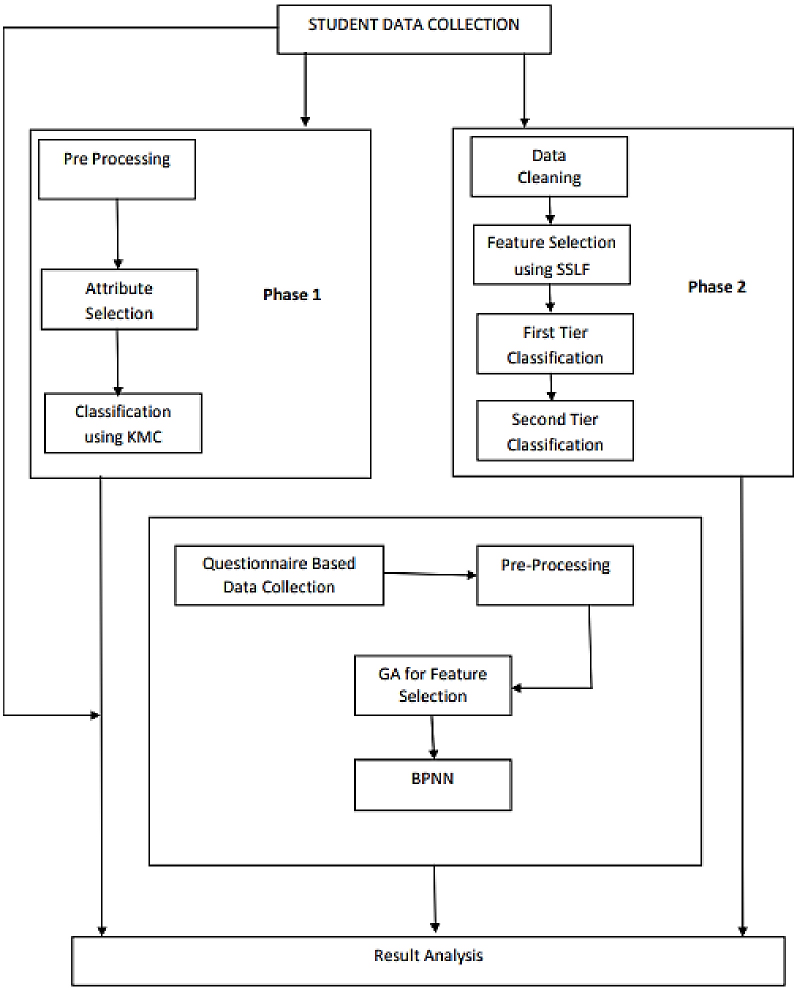

Figure 1: The proposed model

Data Collection

The real time data sets are collected from the student data base of KLN College of Information Technology. The dataset are collected in the questionnaire form, for which the answers are directly obtained from the students. The student variables are used for evaluating the learning and academic activities of students used in the question form that includes school details, attendance, grade points of previous semester. In total, about 350 students are considered for evaluations. The datasets are collected from the first year to final year of students in Computer Science department. In this model, the question set is used for collecting the student details which comprise of learning patterns of students and their academic data. The features for determining the student behaviour used in the question set are student personal data, school data, attendance and grade points. From the data obtained, 250 samples are given for training and 50 are used for testing.

Data Selection and Transformation

In this phase, the data needed for mining the student records are selected. Some features are selected from the database and some of the features are obtained. The data are obtained from the question set and database. First, the attributes selection is processed. The selection process handles the selection of significant attributes for student classification. By evaluating the 24 attributes, the attributes with higher ranks are used for training and testing.

The data features that are required for performing classification method are selected here from the obtained values of databases. The significant features that are selected for classification are listed below.

- Grade points
- Arrear Data
- Attendance
- Entrance Cut-off
- Educational Medium
- Educational Board

Attributes that are considered for feature selection and domain values are presented in Table 1. For accurate classification of students the model considers both the educational data of college as well as school. Based on college data, the classification is done as the Best, Good, Average and Poor. The data from school such as MOE, BT, TM-HS, CGPA, NOA, AP and ECUT are effectively utilized for appropriate classifications. Moreover, the attribute selection process is done with the data from school and college student sample, under various circumstances, with both data samples. The student profile is defined based on their performance and demographic student data.

Table 1: Attributes and domain values

Variables	Description	Domain Values
TM	Total Marks in Higher Secondary	{best, good, average, poor}
MOE	Medium of Education	{Tamil, English}
BT	Board Type	{Matric, StateBoard, CBSE, Diploma}
ECUT	Entrance Cut Off	{good, bad}
CGPA		{1-10}
NOA	Number of Arrears	NA
AP	Attendance Percentage	{50-100}

In this process, for classification, the model uses J48 classifies, which is stated as the model producing efficient classification results. Moreover, the view pattern of the student model is presented as follows,

- Login Data
- Student data
- Student data Evaluations

J48 based Classification

J48 is the tree-based learning model, which is derived from Interative Dichtomiser (ID3). There are some advanced functionalities in J48 model, which includes managing missing domain values, data pruning and so on. Here, WEKA tool is used for the implementation, which is processed with JAVA. Moreover, in this model, the classification operations are processed in consistent manner till the model derives a pure leaf node for producing précised results. The steps are presented below,

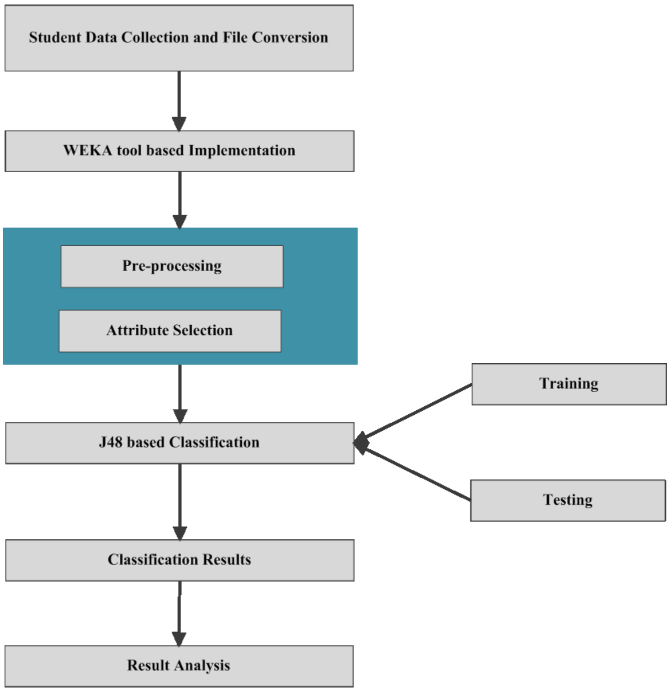

Figure 2: Workflow of the student classification model in Phase 1

Pruning computations

Another significant operation in J48 classifier is pruning. Some features are in the data samples, which are distinctive. The classification functions have been processed with the samples and the tree is constructed. The pruning operations are used here for minimizing the errors in classifications. The computations are given below.

Here, the training student data set is assumed as, $DS = \{ds1, ds2, \ldots, dsn\}$. Each student data instance is given as, $dsi = a1, a2, \ldots$, where, 'a' represents the attributes of 'ds'. Efficient feature selection is processed with this based on the tree node and capable to perform data division based on different classes. The student samples are classified under the class labels are presented in Table 2.

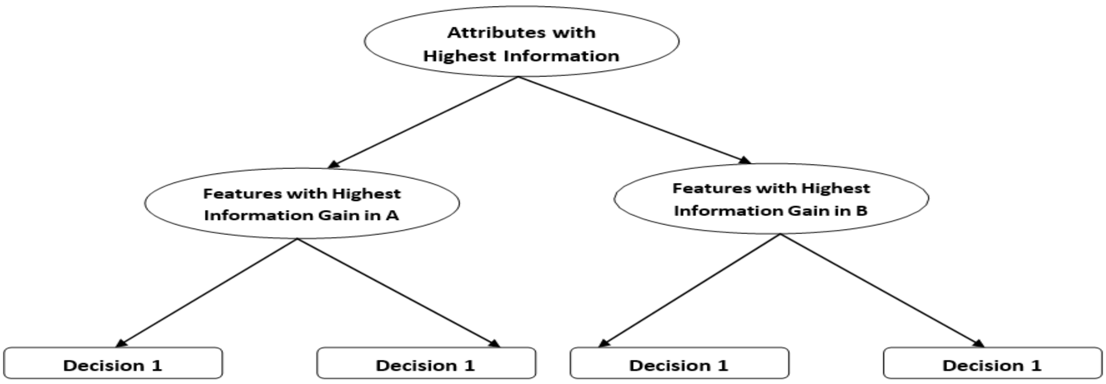

Figure 3: Flow of J48 classification in decision making

Table 2: Samples of KLN data set under classification

AP	CGPA	Result	Class_Labels
More than 90%	80 and above	Pass	**EXCELLENT**
More than 80%	Between 60 and 80	Pass	**GOOD**
More than 85%	Between 45 and 60	Pass	**AVERAGE**
More than 75%	Below 45	Fail	**LOW**

K-Means Clustering (KMC)

KMC is the process of vector quantization, derived from the signal processing techniques that intends to divide 'n' number of samples into 'k' number of clusters. Each sample will come under a cluster, which finds to their nearest mean rate. This may result in dividing the data space into different cells. Moreover, the model reduces within the grouping variations based on the Euclidean distance between the objects. For example, the better solution is derived with the K-median determinations.

Here, the local optimum is derived with the heuristic model, in which the KMC model aims to find the clusters of similar spatial rates.

WEKA Implementation

The major objective of the proposed model is to predict the student performance based on the input variables that are processed and obtained in the model. Here, the Clustering model, built with K-Means Clustering and WEKA tool is used further processing. The datasets are fed into the WEKA tool for operations, which is an open software source that can be processed with large amount of data. The expansion of WEKA is Waikato Environment for Knowledge Analysis. With the obtained data, the student file is converted into Attribute Relation File Format (ARFF) for evaluations in the WEKA tool. Further, the converted student data is opened in console as, '.ARFF' format. The classify panel provides the user to employ the process of classification to the obtained samples, to determine the accurate results and to predict the incorrect results. Here, the classification models such as, Naive Bayes, MLP, J48 and REP Tree. Two testing operations based on the acquired and the trained samples for each classifier are determined. After completing the pre-processing, feature selection is performed with the attribute selection function from WEKA tool, which are having higher rankings for evaluation. As there are no definite dataset for evaluation, it is required to get idea of precision of the developed model. The model provides option to determine the learning rate of students in future examinations. The Design model in WEKA is depicted in Figure 4.

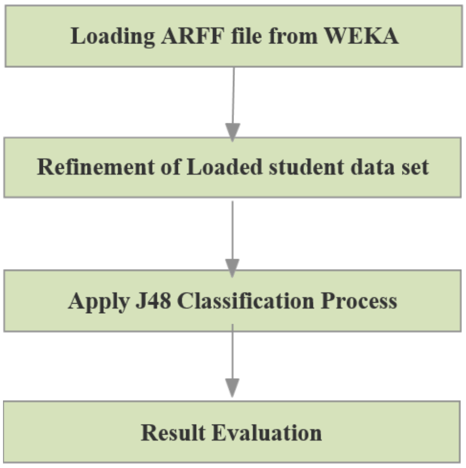

Figure 4: Execution of the design model

RESULTS AND DISCUSSION

The main objective of the proposed work is to detect the student performances in academics based on their records. For implementation, Waikato Environment for Knowledge Analysis (WEKA), which is an open source environment, is used. The student data files are collected and converted into Attribute Relation File Format (ARFF) for data evaluations. Moreover, the results are evaluated based on the factors such as, classification accuracy, precision value and error rates. The results are compared with the existing models such as, Support Vector Machine (SVM), Multi-layer Perceptron (MLP) and Artificial Neural Network (ANN). The classifiers, employed for two testing models are divided for training and testing functions.

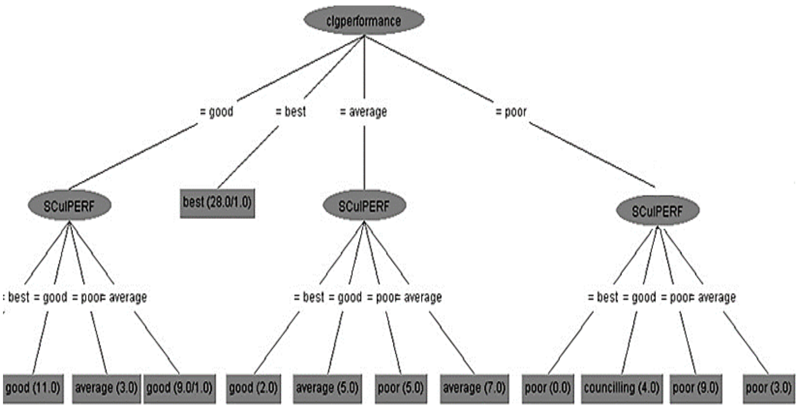

Figure 5: J48 tree representation for student analysis

Model Design

The model design of the classification operation is explained in this section. Among the compared classification model, the J48 algorithm provides better results than other compared models. The modules in the design model comprises of,

1. Log-in Data
2. Student Record
3. Student Result AnalysisJ48 is a tree based classifier model, which is developed based on ID3 algorithm. Moreover, the model uses divide and conquer based node splits, for defining the leaf, child and root nodes. In the given set of 'S' samples, the tree structure is framed as follows

1. If all the samples in 'S' are under the same class or 'S' is having minimal samples, such leaf is considered as the frequent class in 'S'
2. If previous step is not occurred, then selection process is carried out based on single feature with minimal two or more than two possible results. Then, the sample partitions are given as, S1, S2, S3, ..., based on the student cases
3. In recursive manner, the same sets of operations are applied to the other sub nodes.
4. Gain Ratio and Information gain values are ordered based on the heuristic results.

The form results of the proposed model are depicted in the following Figures, in which the log in data are provided in the form in Figure, authentication is in Figure , student data collection is done through the form in Figure, Performance analysis are processed with that. In the results, the student samples are classified under the categories as,

i. Best

ii. Good

iii Average

iv. Poor

The results are given for the tutors to design a new learning methodology based on the classification, thereby improving their results and success ratio, effectively.

The comparative evaluations are presented in the following sections for the parameters classification accuracy, precision values and error rate in Figure 5, Figure 6 and Figure 7, respectively. Their corresponding values obtained on the experimentation in WEKA environment are given in Table 3, Table 4 and Table 5.

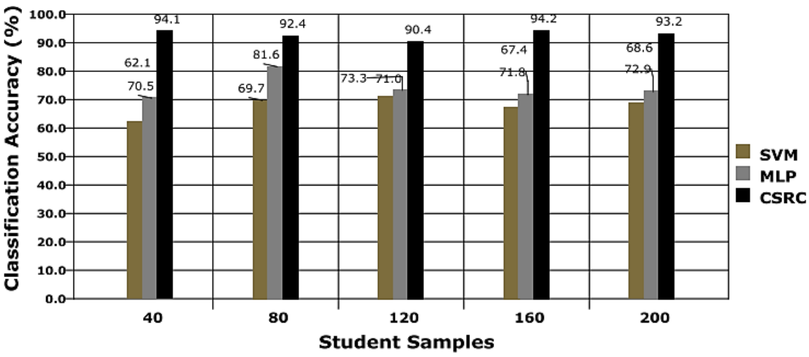

Figure 6: Result comparison for classification accuracy

Table 3: Values obtained for classification accuracy

Models	**Classification Accuracy for Samples**				
	40	**80**	**120**	**160**	**200**
SVM	62.1	69.7	71.0	67.4	68.6
MLP	70.5	81.6	73.3	71.8	72.9
CSRC	94.1	92.4	90.4	94.2	93.2

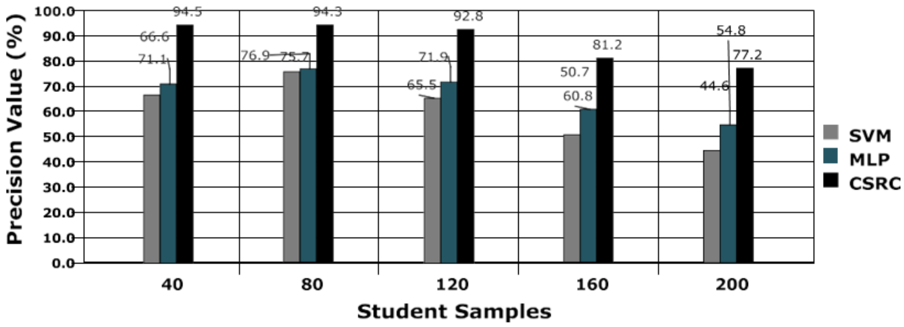

Figure 7: Precision value comparison between models

Table 4: Precision values Vs. Student numbers

Models	**Precision Value for Samples**				
	40	**80**	**120**	**160**	**200**
SVM	66.6	75.7	65.5	50.7	44.6
MLP	71.1	76.9	71.9	60.8	54.8
CSRC	94.5	94.3	92.8	81.2	77.2

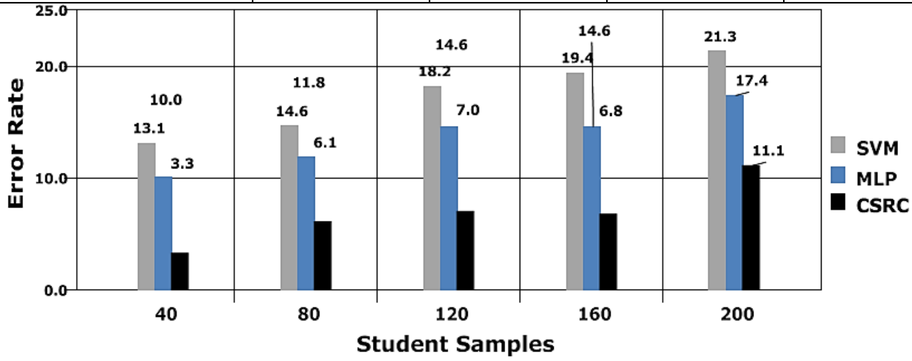

Figure 8: Error rate in classifications

Figure 9: Error rate comparisons

Models	**Error rate for Samples**				
	40	**80**	**120**	**160**	**200**
SVM	13.1	14.6	18.2	19.4	21.3
MLP	10.0	11.8	14.6	14.6	17.4
CSRC	3.3	6.1	7.0	6.8	11.1

The Final comparison is provided here for projecting the classification of students under four class labels, Excellent, Average, Good and Low as mentioned before. The chart is presented in Figure 9. By this proposed model in this phase, the average accuracy rate obtained here is 92.8%, which is further enhanced with machine learning and genetic model incorporations in the next phases of model developments in this research work.

Figure 10: Final results of student classification in SCRC

In this phase, the EDM process involves in evaluating the educational model. The model aims on evaluating the classification accuracy of student performance. The samples comprise of all the personal and academic student parameters. The model accurately provides the student's academic status, to enhance their results in future. The prediction results are not same for all models, which are completely based on the selection of student features effectively. The results analysis shows that the prediction accuracy of compared models is ranged from 80% to 98%. Based on the analysis, it is observed that J48 classifier provides better results than other models and

This research contributes in developing an Accurate Student Classification Model for evaluating the academic performances of students and classifying the input samples. Moreover, researchers in this domain discuss about various types of student behaviours and features that affect the student's result. The EDM techniques refer to several operations of data analysis model with the motive of deriving the hidden knowledge based on the predictive modelling and pattern recognitions. Many higher educational institutions depend on the retention model for reducing the failure rate of students and for increasing their pass percentage. Hence, the proposed work involves in developing a novel classification model by incorporation machine learning methods and data mining models effectively to enhance the student performances in exams and to increase their learning rate, thereby institutional reputation is increased. For developing Accurate Student Classification Model, this work is developed in three-fold manner, as, i. Cluster based Student Record Classification ii. Multi-Tier Student Performance Evaluation Model (MTSPEM) iii. Behaviour based Student Classification System (SCS-B) K-Mean Clustering model (KMC) is used in the first phase of work for student record classification under three classes such as Low performing student, Average Student and Smart Student. The model comprises of the sections, Data Collection and Preparation, Data Selection and Transformation and Student Classification. For performing the classifications, the attributes such as, Grade point, number of arrears, student attendance, Medium of education and Board of education are considered. The model accurately provides the student's academic status, to enhance their results in future. The prediction results are not the same for all models, which are completely based on the selection of student features effectively. The results analysis shows that the prediction accuracy of compared models is ranged from 80% to 98%. Based on the analysis, it is observed that J48 classifier provides better results than other models and the classification is done for student samples as, Excellent, Good, Average and Poor.

REFERENCES

[1]C. Dziuban, C. Graham, P. Moskal, A. Norberg, and N. Sicilia, "Blended learning: the new normal and emerging technologies," International Journal of Educational Technology in Higher Education volume, Feb. 2018.

[2]V. Kovanovic, C. Mazziotti, and J. Lodge, "Learning analytics for primary and secondary schools," Journal of Learning Analytics, vol. 8, no. 2, pp. 1–5, 2021.

[3]S. A. Becker, M. Cummins, A. Davis, A. Freeman, C. G. Hall, and V. Ananthanarayanan, "Nmc horizon report: 2017 higher education edition," The New Media Consortium, Tech. Rep., 2017.

[4]C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," Jan. 2020. doi: 10.1002/widm.1355.

[5]O. Viberg, M. Hatakka, O. Bälter, and A. Mavroudi, "The current landscape of learning analytics in higher education," Computers in Human Behavior, vol. 89, pp. 98–110, 2018.

[6]B. Friedman, P. Kahn, A. Borning, P. Zhang, and D. Galletta, Value Sensitive Design and Information Systems, 01 2006. ISBN 978-94-007- 7843-6

[7](2015) The 2030 agenda for sustainable development. [Online]. Available: https://sdgs.un.org/2030agenda

[8]P. Grimm, "Social desirability bias," Wiley international encyclopedia of marketing, 2010.

[9]K. Raphael, "Recall bias: a proposal for assessment and control," International journal of epidemiology, vol. 16, no. 2, pp. 167–170, 1987.

[10] J. López-Zambrano, J. Lara-Torralbo, and C. Romero, "Early prediction of student learning performance through data mining: A systematic review," Psicothema 2021, vol. 33, pp. 456– 465, 2021. doi: 10.7334/psicothema2021.62.

[11] nalytics," International Journal of Technology Enhanced Learning, vol. 4, pp. 318–331, 01 2012.

[12] A. Peña-Ayala, "Review: Educational data mining: A survey and a data mining-based analysis of recent works," Expert Systems with Applications: An International Journal, vol. 41, pp. 1432–1462, 03 2014.

[13] D. Sansone, "Beyond early warning indicators: High school dropout and machine learning," Oxford Bulletin of Economics and Statistics, vol. 81, no. 2, pp. 456–485, 2019.

[14] M. Barkham, E. Broglia, G. Dufour, M. Fudge, L. Knowles, A. Percy, A. Turner, and C. Williams, “Towards an evidence-base for student wellbeing and mental health: Definitions, developmental transitions and data sets,” Counselling and Psychotherapy Research, vol. 19, no. 4, pp. 351–357, 2019.

[15] R. Palos, L. Maricutoiu, and I. Costea, “Relations between academic performance, student engagement and student burnout: A cross-lagged analysis of a two-wave study,” Studies In Educational Evaluation, vol. 60, pp. 199–204, 03 2019.

[16] D. Murchan and F. Siddiq, “A call to action: a systematic review of ethical and regulatory issues in using process data in educational assessment,” Large-scale Assessments in Education, vol. 9, 12 2021.

[17] J. Reyes, “The skinny on big data in education: Learning analytics simplified,” vol. 59, pp. 75–80, 02 2021.

[18] H. Drachsler and W. Greller, “Privacy and analytics – it’s a delicate issue. a checklist for trusted learning analytics.” 04 2016.

[19] S. Arie, “Simon wessely: “every time we have a mental health awareness week my spirits sink”,” BMJ, 09 2017

[20] N. Sclater, “Developing a code of practice for learning analytics,” Journal of Learning Analytics, vol. 3, pp. 16–42, 04 2016.

[21] B. Chen and H. Zhu, “Towards value-sensitive learning analytics design,” 12 2018.

[22] M. L. Kern, L. E. Waters, A. Adler, and M. A. White, “A multidimensional approach to measuring well-being in students: Application of the perma framework,” The journal of positive psychology, vol. 10, no. 3, pp. 262–271, 2015.

[23] M. E. Seligman, Flourish: A visionary new understanding of happiness and well-being. Simon and Schuster, 2011.

[24] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, Learning from imbalanced data sets. Springer, 2018, vol. 10.

[25] Priyanka and D. Kumar, "Decision tree classifier: a detailed survey," International Journal of Information and Decision Sciences, vol. 12, no. 3, pp. 246–269, 2020.

[26] M. Hajduk, A. Heretik Jr, B. Vaseckova, L. Forgacova, and J. Pecenak, "Prevalence and correlations of depression and anxiety among slovak college students." Bratislavske lekarske listy, vol. 120, no. 9, pp. 695- 698, 2019.