An Intelligent Clustering Technique for Analysing the Performance of Students during Lockdown Period of Covid-19

K.P. Prakash[a], and K Selvakumari[b]

[a]Research Scholar, Department of Mathematics, Vels Institute of Science, Technology, and Advanced Studies, Chennai, Tamil Nadu, India.
[b] Professor, Department of Mathematics,Vels Institute of Science, Technology, and Advanced Studies, Chennai, Tamil Nadu, India.

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

Abstract: Corona virus or simply Corona is the current leading pandemic of the world. It has affected students and their in education in higher numbers than any other sector putting them into a depression. Hence this research attempts to suggest solutions for reducing depression amongst students amidst the pandemic. This work proposes ESVMs (Enhanced Support Vector Machines) model for its predictions. Identifying student performances is complex issue as the numbers are voluminous and hence the objective of this research is to assess student performance prediction model by using an efficient clustering method. Missing values and irrelevant data are resolved in this work using SCCs (Statistical correlation Coefficients) which work on subject wise manner or student wise data. This work also provides a novel solution for data pre-processing. IFCM (Improved Fuzzy C-means clustering) proposed in this work identifies high quality clusters with robustness. Further, the use of PSO (Particle Swarm Optimization) in feature selections improves its efficiency of the given data. Classifications are executed by the proposed ESVMs which predicts student's grade with accuracy. The evaluation results of this study improve classification accuracy significantly when compared to existing prediction models.

Keywords*:*COVID-19, online education, educational data mining (EDM), Statistical correlation Coefficient (SCC), Improved Fuzzy C-means clustering algorithm (IFCM), Particle Swarm Optimization (PSO), Enhanced Support Vector Machine algorithm.

1. Introduction

The pandemic Corona has literally affected mall sectors and millions of people. The education systems in all 192 countries have been stranded by the epidemic. Schools and colleges are closed and almost 1.7 billion students' haven affected academically (Aristovnik, 2020). Most university proceedings stand cancelled or postponed for avoiding the pandemic which rises phenomenally in human gatherings. The pandemic avoidance measures have lead to medical, economical and social implications. On the educational side undergraduates and postgraduates seem to be the worst hit (George, 2020). In an effort to keep education on its toes, most educational institutions have introduces online mode which helps avoid physical contacts between teachers and learners (Tria, 2020). One great issue of online education is the access to resources required for this kind of education. Students suffer in these programmes due to their economical status or digital divide. Across the globe this effect only varies in its degree, but is a pronounced issue. Multitude of factors can be attached to this scenario including age, family background, and access to "substitute" opportunities in education. Thus, closure of educational institutions, pandemic associated health hazards pose major challenges in this sector (Aboagye, 2021).

The education system was caught unawares by the pandemic as they were and unprepared in coping with such a serious situation. Their lags in infrastructures, inefficient teaching methods and ineffective learning were exposed and the institutions failed to guard themselves. This has resulted in weakening student's academic performances along with reduced developmental skills and stunted progress (Toquero, 2020). The pandemic has also widened socio-economic disparities leading to disparities in educational equity. An already average student at school find s it difficult to go up the ladder and instead seems to come down the ladder. Thus, these discrepancies caused by the pandemic in education have to be attended to and normalized (Lestiyanawati, 2020).

Measuring performances of students academic in this time of pandemic is a challenging issue as it is dependent on multitude of factors including socio-economical, psychological, personal and other environmental elements [7]. This study encompasses student performances, detecting risks, student retentions and other student related factors in predictions. This work is aimed innovative projects for schools for improving their reputations and ranking. However, this study limits itself to moderate level schools and institutes (Lederer, 2020). DMTs (Data Mining Techniques) are a great tool in this assessment of student performances.

Studies that classify or predict or infer, consume more effort in extracting significant indicators that form the base for building accurate predictive models (Owusu-Fordjour, 2020). Most feature selections are based on their ranks or selected while learning feature information from datasets, but lesser studies have investigated visualizations or clustering techniques for such selections in analysis (Radha, 2020). Clustering outcomes might help overcome certain hurdles normally found in features selections or extractions. Many studies have reviewed factors that influence student performance prediction models including socio-economical, personal, psychological and other environmental indicators, but with certain limitations (Owusu-Fordjour, 2020). Hence, this work introduces an effective and efficient model for assessing student performances in terms of their marks and the factors influencing their scores. The study stresses on identifying weaker students in the institution so that they can be offered individual assistance by teachers for their enhanced study scores. The proposed work also evaluates accuracy of other classification systems. The rest of the research work is as tails: Section 1 discusses in brief corona's influence on student performances; Section 2 reviews related researches in clustering and MLTs; Section 3 details about this study's proposed prediction model on student performances; Section 4 displays experimental results and Section 5 concludes this thesis with impending work(Khusna, 2020).

2. Literature Review

This section analyses correlated studies that fit into the framework of this study. Several classification algorithms and their performances are also studies for assessing their advantages and disadvantages. The reviews considered are based on educational settings with low education outcomes due to socio-economic factors like Corona pandemic and with the aim of proposing a new prediction model that can evaluate student performances effectively.

DTs (Decision Trees) were used for assessing academic performances in the study of Hamsa et al (Hamsa, 2016). The proposed scheme used DTs and FGs (Fuzzy Genetic) Algorithm. Several parameters alike interior, session, and admittance scores stayed the basis for the study's assessment. Student presence, average scores of session exams and assignments were combined to form Internal marks. Admission scores were compiled as a weighted score which included 10^{th}, 12^{th} and admission marks. In the instance of post graduate scholars, their degree and entrance scores were taken for computing admission scores. The study's model predicted students' performances on each subject which helped teachers identify students who needed improvements. Their early predictions helped improve student performances in their final examinations. Moreover, high performing students could be identified by reputed companies for job recruitments.

ID3 was used to predict academic performances by Altujjar et al in their study. Their scheme used DTs induction in building their prediction model 9 Altujjar, 2016). The data used in the study was generated from IT (Information Technology) academic scores of female bachelors studying second year at King Saud University, Riyadh, Saudi Arabia. The scheme produced reliable predictions on the performance of students which was used by the IT department for student enhancements.

DMTs (Data Mining Techniques) which have the ability to trace relationship between data elements was used by Al-Twijri et al in their study to streamline higher educational institution decisions. The model's outputs assisted strategic decisions of the institutions while regulating institutional student admissions (Al-Twijri, 2015).

A case study on bright students was conducted by Asif et al. Their proposal predicted high performing students at a degree level to help universities focus more on brighter candidates. Students with low academic achievements were also identified (Asif,, 2014). Their data consisted of 347 undergraduate students which was

Cumulative grades of engineers were studied by Adekitan et al. The study's predictive analysis considered 5th year Nigerian University engineering student's CGPA (Cumulative Grade Point Average) were resolute by means of database of study, entry year and the GPA (Grade Point Average) of initial three years as inputs and predicted using KNIME (Konstanz Information Miner) based model. The study took into account six data mining algorithms and obtained an effective accuracy of 89.15% (Miguéis, 2018; Hoffait, 2017; Helal, 2018). The study verified their results with linear and quadratic regression models and recorded R^2 values of 0.955 and 0.957 for the models. The study predicted graduates with poor results or graduates who may not pass, thus helping in early interventions. Automatic MLT was proposed in the study of Zeineddine et al for enhancing prediction accuracies of models which can predict student performance from available data(Zeineddine, 2021).

MLTs were also used by Aluko et al is their study to develop academic success predictions based on previous academic performances (Aluko, 2016). The study used KNNs (K-Nearest Neighbours) and LDA (Linear Discriminant Analysis) where k-NN outperformed LDA model with better accuracy. The study found mathematics grades of even ordinary examinations had significant impact on the undergraduate student's academic successes. The main contribution of this work was in using previous academic performances to evaluate their academic success and thus implying previous academic performance as a useful predictor for judging academic success of students. Further, the study inferred that k-NN based architectures could serve as a valuable tool for academic predictions, especially while selecting new intakes into undergraduate programmes in Nigeria.

MLTs were also proposed by Mengash et al in their study in support of university admission decisions. Their scheme predicted academic performances at the university level where their dataset consisted of 2,039 Computer Science students enrolled over a period of 3 years at a Saudi public university (Mengash, 2020). The scheme's results demonstrated that students initial performance at the university could be predicted with pre-admission standards like school mark averages, SAAT (Scholastic Achievement Admission Test) scores, and General Aptitude Test scores. The study also found that SAAT was enough to predict student's future performances and could be used as the primary factor in academic performance assessment systems. The study's evaluations with ANNs achieved 79% accuracy and performed better than other MLTs including DTs, SVMs and NB.

3. Proposed Methodology

The basic objective of this work is to predict student performances by building an effective clustering and enhancing predictions. Clustering in this proposed methodology groups dataset attributes based on IFCM similarity which is fed into classifiers for better classifications. Clustering identifies and clusters attributes based on their values. Once attribute similarity is executed by IFCM, feature selections are executed as detailed below:

- Initially in this work, missing vales and irrelevant data are resolved by computing SCCs of data in either subject wise manner or student wise manner. This is an excellent pre-processing analogy for evaluation of students.
- IFCM is proposed for cluster identification for better quality and robust groupings.
- PSO is used in feature selection for optimality.
- ESVMs classify the samples to predict accurately student's grade.

This research work evaluates the proposed algorithms with performance measures for their accuracy values and thus selects the most suitable algorithm for creating an efficient student grades classification model. Figure 1 depicts the proposed prediction model for student performance predictions.

4. Data Collection

This research work uses two datasets for evaluating its proposed student prediction model and are detailed below:.

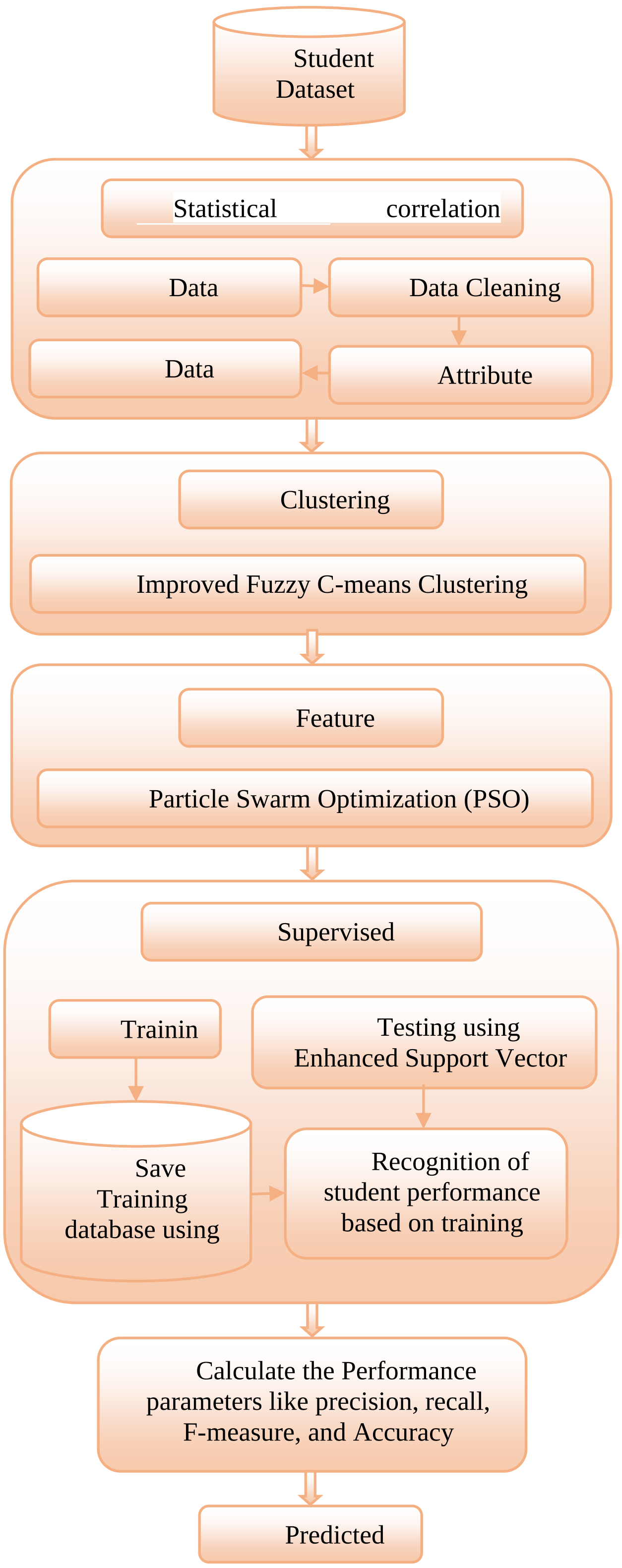
- **First Dataset:** This dataset examines Vietnamese student's in terms of their learning habits when the school is closed due to the pandemic Corona. The dataset SARS-CoV-2 (Covid-19) is an open source to research on potential effects of the pandemic and its prevention (Hoang, 2020). The dataset was generated by consolidating details of questionnaires sent to students using Facebook's educational communities over the period 7-28 February 2020. The questionnaire was divided into three major categories: Individual demographics including family's socio-economic status, type of school and student's occupational aspirations; student learning habits which encompasses their study hours with/without support before and after the pandemic attack and student perceptions on self-learning during the pandemic attack. Though there were 920 views only 460 students responded with answers like born before 2009 or 20 hours of learning per day. These invalid data was not considered and finally the dataset was built from 420 valid observations.
- **Second Dataset:** The second dataset was generated using 1182 students belonging to a mixed age group and their answers for questionnaires. The students belonged to educational institutions in Delhi, the Indian Capital.

4.1. Data Pre-processing

Data pre-processing is the preliminary step of any data processing procedure and transforms collected data into the required format. Incorrect data reduce efficiency and performance of classifiers. This work uses SCC in its data pre-processing.

- **SCC**

Correlations depict degree of associations between variables. It is also the mean of values corresponding to the dependent or outcome or response variable (Asuero, 2006). Assuming a variable $x$is covariate to another variable y, they can be termed as continuous variables (vary together). The strength of this continuity has to be determined for predictions, and hence a statistical assessed correlation coefficient, $r_{xy}(r\ )$, is used for this objective.



Multitude of numerical tools can be used to determine how data fits into this regression. A particularly useful statistical method for this purpose is computing R^2 (depicted as Equation 1) in a linear regression fit. The term is equivalent to the ratio of the sum of squares used in regression (SS_{Reg}) to the total sum of squares of mean deviations (S_{YY}) for a prototypical model using a constant term (homoscedastic case, w_i=1)

$$R^2 = \frac{SS_{Reg}}{S_{YY}} = \frac{S_{YY}-SSE}{S_{YY}} = 1 - \frac{SSE}{S_{YY}} = 1 - \frac{(y_i-\hat{y_i})^2}{(y_i-\bar{y_i})^2} \qquad (1)$$

Where, $\hat{y}$- predicted value of y, $\overline{y}$ - mean of y values, i =1,2, . . . n, *SSE* - residual sum of squares. A prediction model without a constant term computes R^2 as$1-SSE/SST$ where *SST* represents the total of y^2. R^2 in Equation (1) computes total variations in $\overline{y}$ as given by regression. Hence, a greater value of R^2 implies total variations of $\overline{y}$ reduces by the introduction of the independent variable x which is expressed as a percentage. Since $0 \leq SSE \leq SS_{YY}$, it is also true that $0 \leq R^2 \leq 1$. The correlation between y and $\hat{y}(R)$ can be calculated using Rquation (2)

$$R = r_{y\hat{y}} = \frac{\Sigma(y_i - \widehat{y_i})(y_i - \overline{y_i})}{[\Sigma(y_i - \widehat{y_i})]\,[(y_i - \overline{y_i})]^{1/2}} \qquad (2)$$

The result of (2) is generally a multiple correlation coefficient and it it improper to compare R^2 values of different equations obtained from different coefficients derived from the same set of data. Simple regressions with a constant term result in a determined coefficient that equals the square of variable x, y correlation coefficient explained as equation (3)

$$r_{xy} = \pm\sqrt{R^2 = \sqrt{1 - \frac{S_{YY} - a_1^2 S_{XX}}{S_{YY}}}} = a_1\sqrt{\frac{S_{XX}}{S_{YY}}} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} \qquad (3)$$

The regression slope line $a1$ determines the resulting sign, positive + or negative - when R^2 =1, it implies a perfect fit i.e. all points lie on the regression line and if it is zero, it implies y is not a function of x. Regression coefficients are also associated to r_{xy} in case of complex regressions.

Covariance among any two random variables x and y, in a normal joint distribution, measures fluctuations between them and can be defined as anticipated value of the product of the deviations between x and y from their expected values. Covariance can be defined as Equation (4)

$$cov(x,y) = \frac{1}{n-1}\sum(x_i - \bar{x})(y_i - \bar{y}) \qquad (4)$$

$$cov(x,y) \leq s_x s_y \qquad (5)$$

Where the above equations imply $r \leq 1$.

Covariances measure correlations between x and y and outputs a positive/negative value based on their linear relationship. When x and y are autonomous (uncorrelated) covariance value is zero. Though, the converse may not be true as highly dependent non-linear random variables can be constructed with zero covariance. Covariance is a variance and a superior instance of covariance of a random variable with itself. The square root of this variance called SD (Standard Deviation) and denoted by σin a sample population s is always positive. Covariance is normally used in uncertainties.

4.2. IFCM

This work clusters with IFCM used for segmenting images. It divides an image into various clusters based on image pixels value similarity (Pal, 2005). It uses a degree of membership to cluster pixels of an image. It is a recursive clustering technique that produces optimality in partitions by minimizing weights and using a squared error objective function within clusters (Shi, 2001; Xiao, 2014).

Assuming $X = \{x_1, \dots\dots, x_n\}$ is a dataset and integer c>1. X is divided into c clusters or disjoint sets $X_1, \dots., X_c$ such that $X_1 \cup \dots.\cup\ X_c = X$ or equivalent to an indicator function $\mu_1, \dots\dots \mu_c$ such that $\mu_i(x) = 1$. When x is in X_i and $\mu_i(x) = 0$ and if not in X_i, then for all $i = 1, \dots\dots, c$ is dividing X into c clusters $X_1 \dots.. X_c$using $\{\mu_1, \dots\dots \mu_c\}$. A fuzzy function generates $\mu_i(x)$ values in the range [0,1] such that $\sum_{i=1}^{c} \mu_i(x) = 1$ for all x in X. $\{\mu_1, \dots\dots \mu_c\}$ is called a fuzzy c-partition of X and the FCM objective function J_{FCM} can be demarcated as Equation (6)

$$J_{FCM}(\mu, v) = \sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ij}^m d^2(x_j, v_i) \qquad (6)$$

Where, $\mu = \{\mu_1, \dots\dots \mu_c\}$ - fuzzy c-partition and $\mu_{ij} = \mu_i(x_j)$, m – fixed weight exponent greater than one, establishes the degree of fuzziness, $v = \{v_1, \dots\dots v_c\}$ - cluster centers in c, and $d^2(x_j, v_i) = \|x_j - v_i\|^2$ - Euclidean distance. FCM iterates based on conditions for minimizing J_{FCM} and uses updates given by the following equations:

$$v_i = \frac{\sum_{j=1}^{n} \mu_{ij}^m x_j}{\sum_{j=1}^{n} \mu_{ij}^m}\ (i = 1, \dots\dots c) \qquad (7)$$

and

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}(\frac{d(x_j,v_i)}{d(x_j,v_i)})^{2(m-1)}} \quad (8)$$

In each iteration, μ and v get updated based on equations (7) and (8) and the iterations optimize $JFCM(\mu, v)$ until the condition $|\mu(l+1) - \mu^l| \leq \varepsilon$ is satisfied.

Also, from Equation (6), it is clear that FCM objective function does not consider any spatial dependencies in X and considers each data point individually. Further, membership function of (8) is determined by $d^2(x_j, v_i)$, which measures point's similarities with the cluster center. Higher memberships occur when points are closer to the center and hence the membership function is susceptible to noises and artifacts and other noises affect membership degrees resulting in improper segmentations.

- **DCAF (Distance Community Attraction Factor)**

This work overcomes certain drawbacks of clustering by using an improved algorithm. DCAF is computed for neighboring data points. Clustering operations result in points attracting their neighbours and this concept is used DCAF which depends on two factors namely data points or feature attraction$\lambda(0 < \lambda < 1)$, and spatial position of the neighbors or distance $\xi(0 < \xi < 1)$. These factors depend on neighborhood structures. DCAF is depicted as the following equation

$$d^2(x_j, v_i) = \|x_j - v_i\|^2(1 - \lambda H_{ij} - \xi F_{ij}) \quad (9)$$

Where, H_{ij} - feature attractions and F_{ij} - distance attractions. λandξ are parameters that adjust the degree of neighboring attractions. H_{ij}andF_{ij} within a neighbourhood S can be computed using the following Equations:

$$H_{ij} = \frac{\sum_{k=1}^{S}\mu_{ik}g_{ik}}{\sum_{k=1}^{S}g_{ik}} \quad (10)$$

$$F_{ij} = \frac{\sum_{k=1}^{S}\mu_{ik}^2 q_{jk}^2}{\sum_{k=1}^{S}g_{jk}^2} \quad (11)$$

with

$$g_{jk} = |x_j - x_k|,\ q_{jk} = (a_j - a_k)^2 + (b_j - b_k)^2 \quad (12)$$

Where, (a_j, b_j), $(a_k - b_k)$ - point j and k. Higher λ value implies strong feature attractions while higher ξ value implies strong distance attractions and optimization of these two parameters leads to effective segmentations.

4.3. Feature selections with PSO

Features are selected in this work using PSO which use multiple particles. PSOs mimic a swarm moving through a search space looking out for best possible solutions. Each particle represents a point in a D-dimensional space where they fly based on the flying experiences of particles [26]. Flying particle moving with a certain velocity in the D-dimensional space, find optimal solutions. A particle's velocity i can be expressed as $V_i = (v_{i1}, v_{i2}, \ldots, v_{iD})$, while its location can be expressed as$(x, x_{i2}, \ldots, x_{iD})$. The optimal locations for i can be expresses as$p_g = (p_{g1}, p_{g2}, \ldots, p_{gD})$ where global optimum position of particles is g_{best}. Every particle's fitness is computed using a fitness function. PSO's velocity updates in D-dimensional space are depicted in the following Equations (13) and (14):

$$v_{id} = w \times v_{id} + c_1 \times rand() \times (p_{id} - x_{id}) + c_2 \times Rand() \times (p_{gd} - x_{id}) \quad (13)$$

$$(X_{id} = x_{id} + v_{id}) \quad (14)$$

PSO uses many parameters within a population namely Q representing Quantity, w representing inertia weight), acceleration constants C1,C2 = 2, maximum velocity represented by v_{max}, G_{max} representing maximum number of iterations, random functions rand ( ) and rand ( ) generating values in the interval [0,1].

PSO uses these parameters for sharing local/global information which are used for analyzing selected classification parameters in terms of optimization and thus help in testing algorithms using multiple sets of classical functions to verify global search performances of the algorithms.

4.4. Classification using ESVMs

This work proposes ESVMs for modelling its predictions as it can precisely identify performing and non-performing classes of students and thus outline non-performing classes (students).

- **Support Vector Machine**

SVMs have been used in anomaly detections due to their capability to classify non-linear information using a kernel function [27]. This sub-section explains SVMs and then details on the proposed EOC-SVM.

Assuming SVMs support two-class classifications and the set of training instances $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$. $x_i \in R^d$, where n is the samples count, y_i represents the class label of an instance x_i and $y_i \in [-1, +1]$. Linear SVMs classify by maximizing the hyperplane optimally and thus maximize the "margin" of the classifier using $w^T x + b = 0$, where w∈F and b∈R are parameters whitch define the position of the decision hyperplane in a feature space F. The decision function is depicted as Equation (15)

$$f(x, w, b) = sign(w^T x + b) \in \{-1, +1\} \quad (15)$$

Where,

$$sign\ (w^T x + b) = \begin{cases} +1, & if(w^T x + b) \geq 0 \\ -1, & otherwise \end{cases} \quad (16)$$

SVMs were first proposed for linearly separable classifications. SVMs find (w,b) for positioning the hyperplane separating two classes at maximum distance in training samples to minimize errors during generalizations. This distance is defined as margin. SVMs were adopted to classify non-linear data by allowing samples to violate the margin for obtaining non-linear decision boundaries by projecting data into an advanced dimension space and using the non-linear function Φ(x). Data points which cannot be separated linearly are projected into a feature space F for separations. Thus hyperplane when projected back to the input space, it becomes non-linear in shape. SVMs have an issue of over-fitting noisy data; hence slack variables ξ is presented for certain data points to lie surrounded by the margin. The parameter C > 0 is a trade-off between classification errors while training and margin maximizations. SVMs objective function minimization can be defined as Equation (17)

$$\min_{w,b,\xi_i} \frac{\|w\|^2}{2} + C \sum_{i=1}^{n} \xi_i \quad (17)$$

$$\text{subject to } y_i(w^T \phi(x_i) + b \geq 1 - \xi_i$$

$$\xi_i \geq 0, \qquad i = 1, \dots \dots . n$$

Minimization issues are overcome using Lagrange multipliers $\alpha_i, i = 1, \dots \dots n$ and the new decision making rule for a data point x is defined as equation (18)

$$f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b) \quad (18)$$

Each α_i>0 is weighed by the decision making function. As SVMs are rare very few Lagrange multipliers with non-zero values exist. SVM's kernel function is $K(x, xi) = \Phi(x)\ T\Phi(xi)$. The outcomes of decisions rely on vector's dot-product in a feature space F which explicitly projects data points into a higher dimensional space. When the kernel results in same set of values it is called a kernel trick. The kernel may use linear or polynomial or sigmoid functions. This research work uses GRB (Gaussian Radial Base) Function depicted in equation (19)

$$K(x, x_i) = \exp\left(\frac{-\|x - x_i\|^2}{2\sigma^2}\right) \quad (19)$$

Where, σ∈R - kernel parameter and (x, x_i) - measure of dissimilarity. Thus SVms can classify data points into two classes using a non-linear decision function. SVM's kernel functions are powerful and enable SVMs to ptoject data points into an implicit high dimensional feature space without computations on data co-ordinates and simply work with inner products of all data pairs in the feature space. Moreover, SVMs are computationally cheaper when compared to explicit co-ordinate computations.

- **Enhanced OC-SVMs (One-Class SVMs)**

OC-SVMs separate data into a specific target class and are trained with only positive samples from the target class. OC-SVMs separate data points from the origin by maximizing the distance of the hyperplane in a feature space resulting in a binary function that captures input space regions. The function returns +1 for small captured regions while for other regions it returns -1

$$\min_{w,\xi_i,\rho} \frac{\|w\|^2}{2} + \frac{1}{\eta n} \sum_{i=1}^{n} \xi_i - \rho \quad (20)$$

$$\text{subject to } (w, \phi(x_i) \geq \rho - \xi_i$$

$$\xi_i \geq 0, \qquad i = 1, \ldots \ldots . n$$

η is used as the regularization parameter instead of C where the range of C is from zero to infinity, but the range of η is limited to the interval [0, 1] which helps interpretable solutions where : training samples above the upper bound are regarded as out-of-class and values within the lower bound are used as support vectors. Lagrange techniques and kernel dot-product calculations, change the decision function into:

$$f(x) = sign(\left(w\phi(x_i)\right) - \rho) \qquad (21)$$

$$= sign(\textstyle\sum_{i=1}^{n} \alpha_i K\,(x, x_i) - \rho) \qquad (22)$$

OC-SVMs thus creates a hyperplane with w and ρ which are maximally distanced from the feature space's origin and thus separating all data points from the origin.

- **Adaptive function based One Class-Support Vector Machine (OC-SVM)**

OC-SVM hyper-parameters are automatically fit. In OC-SVM, the kernel parameters γ and regularization parameter η are chosen and (γ_i,η_j) is the learning configuration. Hence, OC-SVM is run with several learning configurations and the best configuration is selected by evaluations of the adaptive function. One significant improvement done for OC-SVMs in this study is the slack variables. Non-zero slack variable ξi permits a point xi to lie on the other side of the decision margin as shown in Figure 2. Robust OC-SVMs use slack variables proportional to the distance from the centroid as it consents points distant from the center to have a huge slack variable values. As slack variables are constants, they are eliminated in minimization objectives. Alternatively, it shifts the decision boundary towards normal points. A portion of the interpretability is lost in results as unrestricted count of data points could perform on the former side of the decision edge.

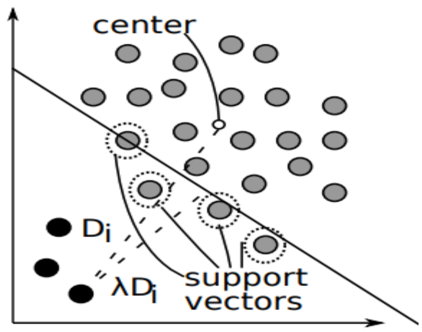


Figure 2. Modifying the slack variables for enhanced one-class SVMs.

Figure 2.illustrates changes in the slack variables. Points that are far from the centroid have a larger slack value and hence the decision boundary shifts nearer to normal points and outliers cease to be support vectors. The objective of this work's proposed EOC-SVMs (Enhanced OC-SVMs), slack variables are not used for minimization objectives and are treated as constraints as Dˆi. Q is an adaptive function.

$$min_{w,p} \frac{\|w\|^2}{2} - \rho \qquad (23)$$

subject to $w^T \phi(x_i) \geq \rho - \lambda * \widehat{D_i}$

slack variable $\widehat{D_i}$ is calculated and represents centroid distance in the kernel space. As Q, the adaptive function is obliquely distinct by the kernel it cannot be used directly. The approximations are summarized, $\frac{1}{n}\sum_{i=1}^{n} \phi(x_i)$is a constant to be released while the normalized distance $\widehat{D_i}$ is used in the optimization Objective.

$$D_i = \left\| \phi(x_i) - \frac{1}{n}\textstyle\sum_{i=1}^{n} \phi(x_i) \right\|^2 \qquad (24)$$

$$\widehat{D_i} = \frac{D_i}{D_{max}} \qquad (25)$$

$$\approx Q(x_i, x_i) - \frac{2}{n}\textstyle\sum_{j=1}^{n} Q(x_i, x_j) \qquad (26)$$

The twin objective of the enhanced one-class SVM can be abridged as tracks:

$$min_\alpha \frac{\alpha^T Q^\alpha}{2} + \lambda D^T \alpha \qquad (27)$$

subject to $0 \leq \alpha \leq 1, e^T\alpha = 1$ (28)

Thus, it is evident from the proposed modification that the dual objective of EOC-SVMs in Equation 26 is obtained and can be incorporated easily.

5.Results and Discussion

The entire model has been designed using MATLAB as a simulator, and the performance has been measured for different student's records. A supervised data classification method was used to regulate the finest prophecy model that apt the necessities for providing an ideal outcome. The performance of the proposed ESVM based classification systems was determined in the work; performance was centred on the four customary estimation metrics for accuracy, sensitivity, specificity and f-measure. The admissions in the confusion matrix have the subsequent significance in the perspective of a data mining problem: a is the correct negative prediction, also termed true negative (TN), categorized as failed by the model; b is the incorrect positive prediction, also called false positive (FP), classified as passed by the model; c is the incorrect negative prediction, also named false negative (FN), classified as failed by the model; and d is the correct positive prediction, also called true positive (TP), classified as passed by the model. The performance metrics conferring to this confusion matrix are considered as follows.

- **Accuracy**

The accuracy (AC) is the proportion of the total number of predictions that were correct. It is dogged using the resulting equation:

$$AC = (d + a)/(d + a + b + c) \tag{29}$$

- **Sensitivity**

The recall or TP rate is the proportion of positive cases that were correctly recognized, as designed using the following equation:

$$Sensitivity = d/(d + c) \tag{30}$$

- **Specificity**

The TN rate is the proportion of negatives cases that were correctly classified as negative, as calculated using the following equation:

$$Specificity = \frac{a}{a + b} \tag{31}$$

- **F-Measure**

The confusion matrix fits to a binary classification, recurring a value of either "passed" or "failed". The sensitivity and specificity actions might clue to biased explanations in the assessment of the model, as intended using the subsequent equation:

$$F - measure = \frac{2d}{2d + 2b + c} \tag{32}$$

Table .1. Performance comparison table for proposed and existing student prediction model

Methods	Accuracy	Precision	Recall	F-measure	Specificity	Error rate
OFGD	84	75	79	76.948	80	16
OTL	88.579	78.969	89.401	83.862	87.586	11.421
ESVM	91.892	81.983	91.682	86.562	90.153	8.1081
IFCM-ESVM	93.187	83.401	92.892	87.891	93.102	6.8127

Table .1.illustrate the performance comparison results for proposed and existing student prediction model.

The classification accuracy of the prediction model has been examined, as shown in Figure 3. The average accuracy computed for the designed student performance prediction model using Online Focus Group Discussion (OFGD), Online Teaching Learning (OTL) Enhanced Support Vector Machine (ESVM) and Improved Fuzzy C

Mmeans clustering (IFCM) with Enhanced Support Vector Machine (ESVM) is 84%, 88.57%, 91.89% and 93.18% respectively

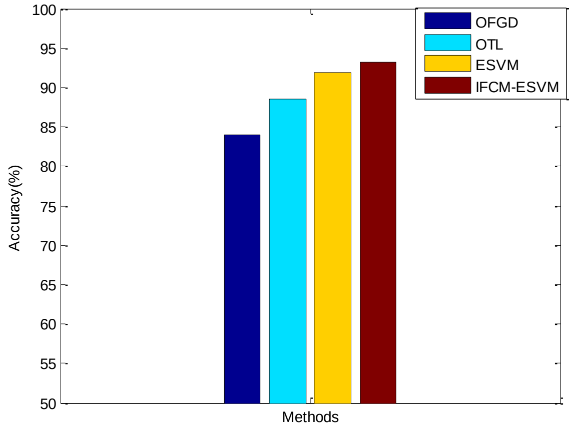


Figure.3. Accuracy comparison between the proposed and existing student prediction model

.

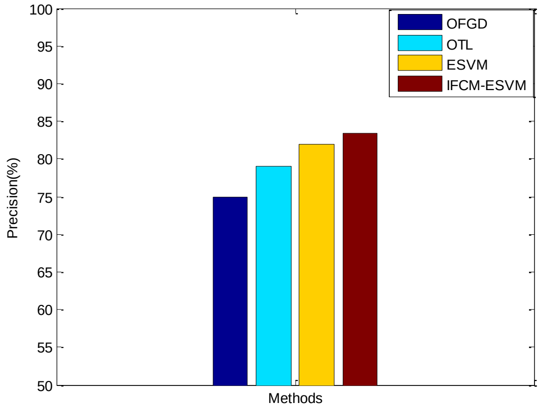


Figure.4. Precision comparison between the proposed and existing student prediction model

Figure 4.represents the precision parameters, which show the correctly classified student data compared to the total number of data that is (misclassified and accurately classified). From the graph, it is clear that using ESVM-IFCM the classification rate of accurately classified student cases is high compared to an existing prediction model.

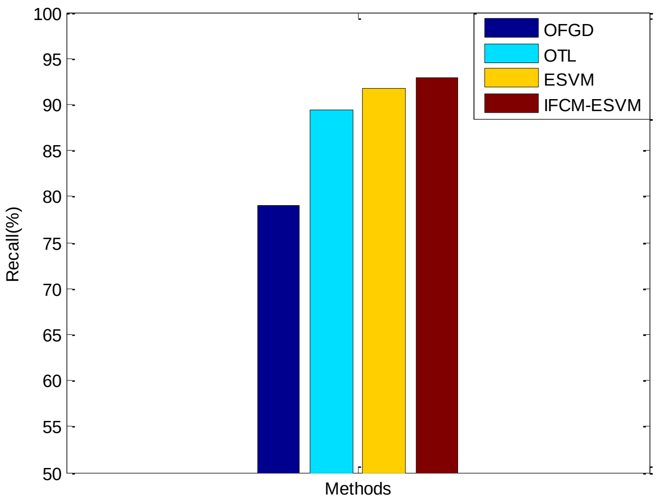


Figure.5. Recall comparison between the proposed and existing student prediction model

Figure 5 illustrated the recall values analyzed by varying the total number of student's records as represented along the x-axis. The recall represents the rate of correctly classified student cases with respect to the total number of correctly and unclassified students records. From the results it is concluded that the proposed ESVM-IFCM prediction model has recall values compare to the existing prediction models.

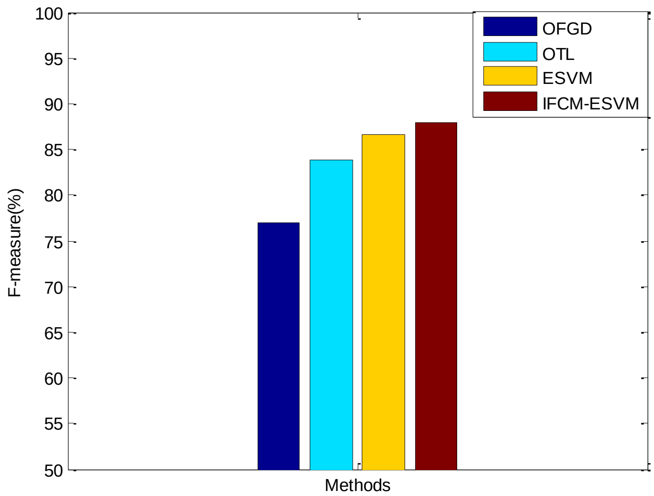


Figure.6. F-measure comparison between the proposed and existing student prediction model

As shown in figure .6, the F- Measure is used to show the relationship between the observed precision and recall values. The average value of F-measured analyzed for Online Focus Group Discussion (OFGD), Online Teaching Learning (OTL), Enhanced Support Vector Machine (ESVM) and Enhanced Support Vector Machine (ESVM-IFCM) is 76.94%, 83.86%, 86.56% and 87.89% respectively.

Table 2. Dataset Comparison

	Accuracy	**Precision**	**Recall**	**F-measure**	**Specificity**	**Error rate**
Dataset 1	93.187	83.401	92.892	87.891	93.102	6.8127
Dataset 2	95	88.88	95.239	91.949	86.723	5

Table 2.tabulates the performance analysis of the proposed student prediction model for both dataset. The dataset 1 has less number of student records (430) and the dataset 2 has more number of student records (1180).

From this, it is clearly identified that the proposed model achieves high accuracy as 95% for dataset 2 and also 93.187% for dataset 2. The proposed student prediction model provides highest accuracy where as the student records are increased.

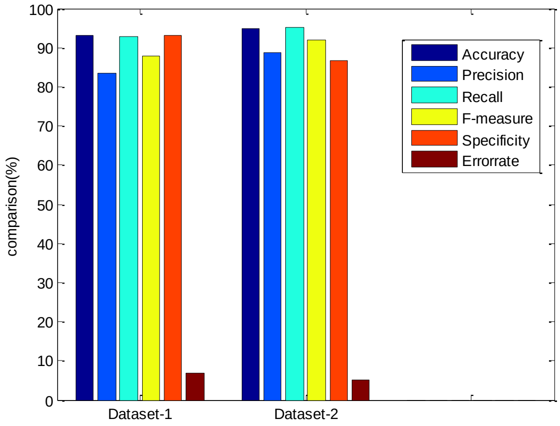


Figure .7. Performance comparison for two datasets

The figure .7.shows the performance comparison results for two datasets. In the dataset 1 has 430 student records where as the dataset 2 has 1180 student records. From the results, it concludes that the proposed student prediction model has high accuracy when the student records are increased.

6. Conclusion

The COVID-19 epidemic is ascertaining to be a productive disruptor, providing a prospect for restructuring the current conventional, classroom based edifying system. The rapid evolutions to online mode supported in keeping permanence of optometry education programs, effectually fitting in the drive of accomplishment of the existing academic year. The speedy transition to online education has not only promoted optometry students but also has fashioned a momentum of continual education for committed optometrist in the country. This research work utilize the data mining systems which permit a high level abstraction of data from fresh data, posing stimulating potentials for the education field. This research work presented the student performance prediction model by proposing the efficient clustering method for improving the performance of the prediction model. As a result, having the info created through this proposed research, institute would be able to recognise students at risk early, and afford enhanced additional training for the weak students. Then, it seems that data mining has a lot of prospective for education. Further this work extended as hybrid learning for increasing the classifier performance.

References

1. Aristovnik, A., Keržič, D., Ravšelj, D., Tomaževič, N., & Umek, L. (2020). *Impacts of the COVID-19 pandemic on life of higher education students: A global perspective. Sustainability, 12(20), 8438.*
2. Aboagye, E., Yawson, J. A., & Appiah, K. N. (2021). COVID-19 and E-learning: The challenges of students in tertiary institutions. *Social Education Research*, 1-8.
3. Altujjar, Y., Altamimi, W., Al-Turaiki, I., & Al-Razgan, M. (2016). Predicting critical courses affecting students performance: a case study. *Procedia Computer Science, 82,* 65-71.
4. Al-Twijri, M. I., & Noaman, A. Y. (2015). A new data mining model adopted for higher institutions. *Procedia Computer Science*, *65*, 836-844.
5. Asif, R., Merceron, A., & Pathan, M. K. (2014).Predicting student academic performance at degree level: a case study. *International Journal of Intelligent Systems and Applications*, *7*(1), 49.
6. Adekitan, A. I., & Salau, O. (2019).The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon*, *5*(2), e01250.
7. Asuero, A. G., Sayago, A., & Gonzalez, A. G. (2006). The correlation coefficient: An overview. *Critical reviews in analytical chemistry*, *36*(1), 41-59.
8. Aluko, R. O., Adenuga, O. A., Kukoyi, P. O., Soyingbe, A. A., & Oyedeji, J. O. (2016). Predicting the academic success of architecture students by pre-enrolment requirement: using machine-learning techniques. *Construction Economics and Building*, *16*(4), 86.

9. Arunkarthikeyan, K., Balamurugan, K., Nithya, M. and Jayanthiladevi, A., 2019, December. Study on Deep Cryogenic Treated-Tempered WC-CO insert in turning of AISI 1040 steel. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) (pp. 660-663). IEEE.
10. Balamurugan, K., Uthayakumar, M., Ramakrishna, M. and Pillai, U.T.S., 2020. Air jet Erosion studies on mg/SiC composite. Silicon, 12(2), pp.413-423.
11. Balamurugan, K., 2020. Compressive Property Examination on Poly Lactic Acid-Copper Composite Filament in Fused Deposition Model–A Green Manufacturing Process. Journal of Green Engineering, 10, pp.843-852.
12. Deepthi, T., Balamurugan, K. and Balamurugan, P., 2020, December. Parametric Studies of Abrasive Waterjet Machining parameters on Al/LaPO4 using Response Surface Method. In IOP Conference Series: Materials Science and Engineering (Vol. 988, No. 1, p. 012018). IOP Publishing.
13. George, M. L. (2020). Effective teaching and examination strategies for undergraduate learning during COVID-19 school restrictions. *Journal of Educational Technology Systems*, *49*(1), 23-48.
14. Hamsa, H., Indiradevi, S., & Kizhakkethottam, J. J. (2016). Student academic performance prediction model using decision tree and fuzzy genetic algorithm. *Procedia Technology*, *25*, 326-332.
15. Hoffait, A. S., & Schyns, M. (2017).Early detection of university students with potential difficulties. *Decision Support Systems*, *101*, 1-11.
16. Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D. J., & Long, Q. (2018).Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*, *161*, 134-146.
17. Hoang, A. D., Nguyen, Y. C., Dinh, V. H., & Pham, H. H. (2020).Dataset of Vietnamese Student's Learning Habit during School Closure due to COVID-19 Pandemic. *Mendeley Data*.
18. Khusna, F. A., & Khoiruddin, M. (2020, October). SURVIVING ONLINE LEARNING CHALLENGES DURING THE COVID-19 PANDEMIC, CAN WE?.In *Language and Language Teaching Conference 2020*.
19. Latchoumi, T.P., Dayanika, J. and Archana, G., 2021. A Comparative Study of Machine Learning Algorithms using Quick-Witted Diabetic Prevention. Annals of the Romanian Society for Cell Biology, pp.4249-4259.
20. Lestiyanawati, R. (2020). The Strategies and Problems Faced by Indonesian Teachers in Conducting e-learning during COVID-19 Outbreak. *CLLIENT (Culture, Literature, Linguistics, English Teaching)*, *2*(1), 71-82.
21. Lederer, A. M., Hoban, M. T., Lipson, S. K., Zhou, S., & Eisenberg, D. (2020).More than inconvenienced: the unique needs of US college students during the CoViD-19 pandemic. *Health Education & Behavior*, 1090198120969372.
22. Mengash, H. A. (2020). Using data mining techniques to predict student performance to support decision making in university admission systems. *IEEE Access*, *8*, 55462-55470.
23. Owusu-Fordjour, C., Koomson, C. K., & Hanson, D. (2020).The impact of Covid-19 on learning-the perspective of the Ghanaian student. *European Journal of Education Studies*.
24. Miguéis, V. L., Freitas, A., Garcia, P. J., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, *115*, 36-51.
25. Owusu-Fordjour, C., Koomson, C. K., & Hanson, D. (2020).The impact of Covid-19 on learning-the perspective of the Ghanaian student. *European Journal of Education Studies*.
26. Pal, N. R., Pal, K., Keller, J. M., & Bezdek, J. C. (2005).A possibilistic fuzzy c-means clustering algorithm. *IEEE transactions on fuzzy systems*, *13*(4), 517-530.
27. Radha, R., Mahalakshmi, K., Kumar, V. S., & Saravanakumar, A. R. (2020). E-Learning during lockdown of Covid-19 pandemic: A global perspective. *International journal of control and automation*, *13*(4), 1088-1099.
28. Ranjeeth, S., Latchoumi, T.P., Sivaram, M., Jayanthiladevi, A. and Kumar, T.S., 2019, December. Predicting Student Performance with ANNQ3H: A Case Study in Secondary Education. In *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* (pp. 603-607). IEEE.
29. Shi, Y. (2001, May). Particle swarm optimization: developments, applications and resources. In *Proceedings of the 2001 congress on evolutionary computation (IEEE Cat. No. 01TH8546)* (Vol. 1, pp. 81-86).IEEE.

30. Tria, J. Z. (2020). The COVID-19 pandemic through the lens of education in the Philippines: The new normal. *International Journal of Pedagogical Development and Lifelong Learning*, *1*(1), 2-4.
31. Toquero, C. M. (2020). Challenges and Opportunities for Higher Education Amid the COVID-19 Pandemic: The Philippine Context. *Pedagogical Research*, *5*(4).
32. Zeineddine, H., Braendle, U., & Farah, A. (2021). Enhancing prediction of student success: Automated machine learning approach. *Computers & Electrical Engineering*, *89*, 106903.
33. Yookesh, T.L., Boobalan, E.D. and Latchoumi, T.P., 2020, March. Variational Iteration Method to Deal with Time Delay Differential Equations under Uncertainty Conditions. In 2020 International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 252-256). IEEE.
34. Xiao, Y., Wang, H., & Xu, W. (2014).Parameter selection of Gaussian kernel for one-class SVM. IEEE transactions on cybernetics, 45(5), 941-953.