J. Electrical Systems 20-9s (2024): 2222-2231

[1,2] Yao Yu

[2,3] Shengfu Wang*

Research on University Education Management Strategy Supported by Artificial Intelligence Big Data Technology

Abstract: - Educational data mining is a practical method for unearthing the connections hidden in educational data and predicting students' academic progress. The findings of the midterm exam grades are the main data in this study's innovative machine learning-based model that forecasts undergraduate students' final exam scores. This study suggests a unique method for analysing the academic performance of distance learners utilising computer technology and machine learning approaches. Here, academic performance of distance education students is the input data that is gathered and processed for noise reduction, normalisation, and smoothing. Then, using a spatio-markov Gaussian model (SMG) and a fuzzy Q-bayes gradient vector neural network (FQBGVNN), the characteristics of the processed data were retrieved. For various student performance analysis datasets, experimental analysis is done in terms of training accuracy, average precision, recall, and MSE. In terms of how long each algorithm takes to produce the results, a comparison of the two methods is also done.

Keywords: Distance Education, Students, Academic Performance, Computer Technology, Machine Learning, Gaussian Model, Fuzzy Model.

1. Introduction:

Schooling is one of the basic parts in the improvement of the country. Somewhat recently, there is a movement in the instructive field like improvement of electronic instructive applications, learning the executives framework, mechanized evaluation frameworks, and gamification in learning [1]. A few robots have been produced for showing small kids. These educational robots have made learning much better and made activity-based learning more popular. Scientists and professionals concur that such procedures are best whenever applied in an understudy's most memorable year of study. Thus, a ton of spotlight has been put on foreseeing, as soon as could really be expected, weak understudies who are inclined to drop their courses [2]. As of late, prescient investigation has depended on AI to help business navigation. Applications in money, tasks and hazard the executives are great authentications of the importance of AI research in different business capabilities. Numerous enhancements have been made in training utilizing information mining methods. Information mining is a method of extricating concealed designs from enormous data sets. The fields of marketing, real estate, customer relationship management, engineering, web mining, and other related fields can all benefit from employing these data mining concepts and techniques. The emerging field of educational data mining is concerned with creating and improving methods for extracting knowledge from educational sector data [3]. The data, which primarily includes the student's personal information and academic performance, can be gleaned from previous experiences or operational data stored in educational institution databases. Additionally, the data can be obtained from e-learning database systems, which contain a significant amount of data and information utilized by the majority of institutions. Numerous strategies have been utilized for appropriate execution of information mining ideas like Guileless Bayes and KNN. Utilization of these strategies gives various types of information which can be found utilizing grouping and bunching. By utilizing these we can extricate information that depicts understudy's exhibition in assessment and all their itemized data. Cluster analysis is used to classify the raw data from the large amount of data that was obtained [4]. A group of physical or abstract objects that are divided and regrouped together based on their similarity is known as clustering. Under the umbrella of AI, machine learning could be a subfield. AI makes progress toward appreciating the intricacy of grouped kinds of gathered information and recognizing the right model for the information by attempting a few models. This can be actually systemized with more straightforward understanding and use by individuals. Despite being a branch of engineering science, machine learning is not the same as the fundamental computing techniques that are employed to solve issues. Machine learning algorithms are developed in a way that enables a machine or system to take input data, generate training sets, and employ statistical estimation to produce the desired range of output [5]. The ideal resource for various universities is an understudy. Colleges and students have a significant role in producing alumni of the highest calibre through their achievements in scholarly exposition. In any case, instructive presentation accomplishment shifts for various reasons understudies might

[1] Nanjing City Vocational College , Nanjing City,Jiangsu Province,21000, China
[2] Namseoul University , Cheonan ,Chungcheongnam-do,31020, Korea
[3] Linyi Vocational University of Science and Technology , Linyi City,Shandong Province,27600, China
Email：1xiayu_0819@sina.com, 2*wangshengfu2024@163.com
corresponding author: Shengfu Wang

have at various degrees of execution accomplishment. In every one of the essential pieces of an understudy's private and expert improvement is execution assessment. In performance evaluations, students' strengths and weaknesses are emphasized. This is an essential tool for building on their strengths and identifying goals for areas that need improvement. Teachers can focus on the essentials, advise and direct students in the right direction, and recognize and reward their accomplishments by conducting research on their students' performance [6].

2. Related works:

AI is utilized in the field of advanced education the executives. According to a poll done in the work of [7], there has been a rise in interest in using ML to forecast student performance and spot pupils at risk using initial data obtained during their years of study. Less work tended to the forecast of understudy execution utilizing information preceding beginning their scholarly excursion. Comparable to distinguishing basic credits, a few variables might influence an understudy's exhibition like social and financial standing, mental components, socioeconomics, educational systems, and interpersonal organizations [8]. Audits of the normal credits utilized in foreseeing understudy execution examined a few factors and sorted them as either inner or outside [9]. Internal assessment includes things like attendance, quizzes, class tests, and assignment marks [10]. A few papers have likewise utilized combined grade point normal (CGPA) as their super inward traits to survey understudy execution [11]. As far as outside appraisal, one necessities to specify understudy socioeconomics, for example, orientation, age, family foundation, extraordinary requirements, and so on [12]. Socio-demographic characteristics, extracurricular activities, high school background, and social interaction network are also popular external attributes [13]. A few scientists have likewise utilized psychometric factors, for example, individual interest, concentrate on propensities, and family support. Work [14] conducted a comprehensive literature review to examine the learning analytics and deep learning studies' evidence. They distinguished understudy execution [15], understudy appraisal and hand composing acknowledgment, just like a portion of the areas where profound learning was conveyed, demonstrating better compared to the pattern models. Various models have been investigated in the learning examination research worldview, but distinguishing the meaning of profound learning in the learning examination space is still in its earliest stages, with concentrates on the reception of this strategy arising over the most recent couple of years [16]. Learning examination contains a few features with the consideration of social event, gathering, looking at and breaking down understudies' data to improve comprehension of the learning climate, bringing about upgraded students' and educators' exhibition. Work [17] examined the need of learning examination applications, customized for each course, to have a superior comprehension of a students discernment, teaching method aim and online way of behaving. Learning examination in versatile learning is a generally new developing region, forming strategies for surveying the way of behaving of portable students. Multi-facet Discernment showed the best precision of 75% for the dataset. Elementary school execution expectations were made by [18] of understudy scholarly execution utilizing past test results. Scientists utilized choice tree, Innocent Bayes, and Zero R characterization calculations. work [19] proposes a model to foresee the last grades of understudies in the class. Ideal expectation is advanced by the calculation on the web, as well as the ideal chance to give the forecast. These are based on how students have done in the past. The technique was demonstrated to be powerful in creating opportune forecasts to empower convenient mediation by the teacher. Creator [20] utilized a web-based calculation to foresee grades for every understudy in a specific class. The calculation they involved expansions in exactness throughout the span of the semester while utilizing just the class information. The calculation additionally shows more prominent exactness than straight relapse and choice tree while contrasting models utilizing similar informational index.

3. System model:

This section provides a thorough explanation of dataset, pre-processing methods, and ML methods used in the study. There are 32 total properties in the two informative collections. Additionally, there are three grades in the informational index: Grades 1, 2, and 3. The student's first period grade is grade 1, their second period grade is grade 2, and their final period grade is grade 3. A pupil is deemed successful if they get an overall grade of 10 or better. For numerical features, the terms mean and standard deviation are employed, whereas count is utilized for categorical features. C addresses undiluted information in the table below, and N addresses numeric datatype. Set includes information on demographics, family structure and status, way of life, educational information, and academic performance of students. Class name includes the words "pass" and "come up short." Figure 1 shows the suggested model.

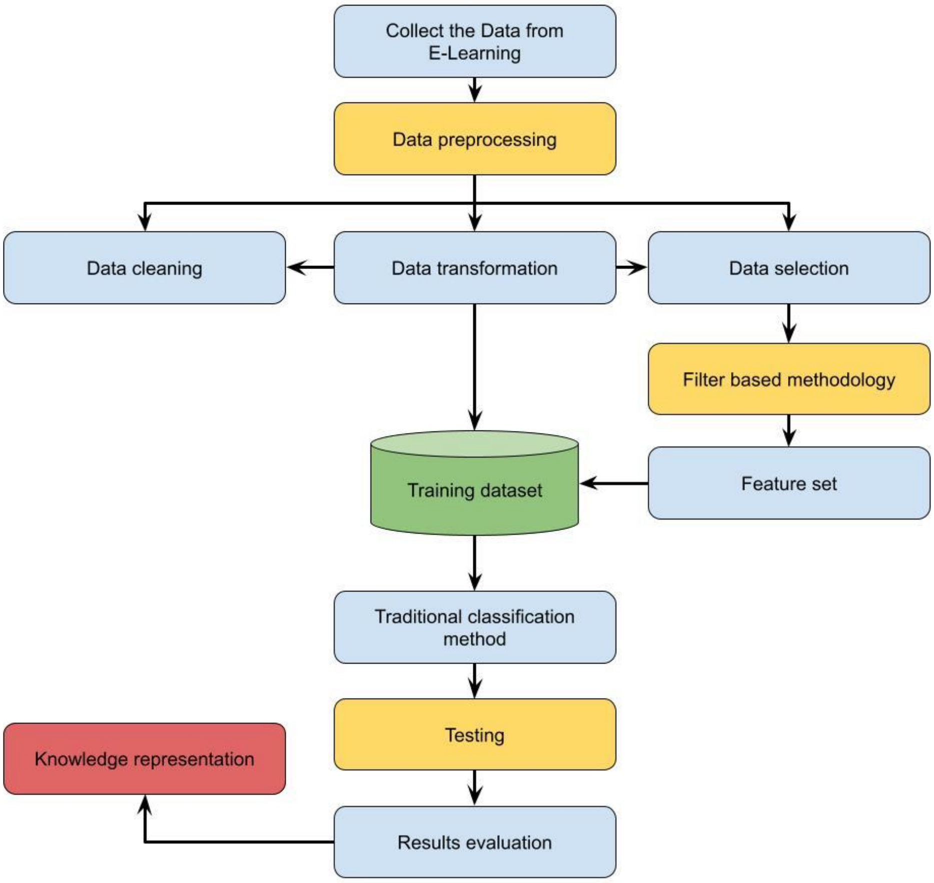

Figure 1- proposed distance education based students academic performance analysis

Binary user-course matrix Y ∈ R |U|×|V| | and sets of users and courses U and V should be defined. If user u has a record of taking the course v, then Y = 1; otherwise, Y = 0, where Y denotes whether user u takes course v. Purpose of implicit feedback recommendations is to create a course list based on user's preferences as determined by equation (1).

$$Y_{uv} = \begin{cases} 1, & \text{if user } u \text{ takes course } v \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The online platform should take into account extra data such as subject, job, and language skills when recommending a personalised course to consumers. As was already mentioned, more details can enhance the effectiveness of recommendations. So, we define the dataset that was gathered as follows. Let's say a set of user behaviour features is provided to us. U = (U, J, C, L), where U = n u1, n j1, n j2, , u |U| o indicates set of users, C = n c1, n c2, , c|C| o indicates set of certificates, and L = n l1, l2 , l |L| o indicates set of language skills. Here, the letters |U|, |J|, |C|, and |L| stand for respective quantities of users, occupations, certificates, and language proficiency. U denotes a collection of user behaviour characteristics, where each characteristic is denoted by a single instance (u, j, c, and l), where u is a user, j denotes a job, c denotes a certificate, and l denotes language proficiency. Course attribute feature is defined as C = (V, M), where V = n v1, n v2, , v|V| o denotes set of courses and M = n m1, n m2, , m|M| o denotes the set of subjects.

Spatio markov Gaussian model:

Assume that Xt represents a realisation of the MRF. By optimising the posterior probability distribution provided in equations (2, 3), the label field is assessed for any feature extraction problems.

$$\hat{x}_t = \arg\max_{x_t} P(X_t = x_t \mid Y_t = y_t) \quad (2)$$

$$\hat{x}_t = \arg\max_{x_t} \frac{P(Y_t=y_t|X_t=x_t)P(X_t=x_t)}{P(Y_t=y_t)} \quad (3)$$

Assume that the MRF that $\hat{x}_t$ represents is realised in Xt. The posterior probability distribution presented in equations (4) can be maximised to determine the label field for any feature extraction problem.

$$E(|y_t \cap A_1||\gamma_{t+\Delta t} \cap A_2|) = \int_{A_3}\int_{A_t} k_{t,\perp}(x_1,x_2)dx_2\, \mathrm{d}x_1 + \int_{A_1\cap A_s} k_{t,L}(x)\mathrm{d}x,\ \frac{\mathrm{d}}{\mathrm{d}t}k_t(\eta) = (L^{\Delta}k_t)(\eta), \quad (4)$$

a differential equation with the form eq. (5)

$$\frac{\mathrm{d}}{\mathrm{d}t}k_t^{(1)}(x) = -mk_2^{(\mathrm{d})}(x) - \int_{R^t} a^-(x-y)k_t^{(2)}(x,y)dy + \int_{R^t} a^+(x-y)k_t^{(1)}(y)dy. \quad (5)$$

eq. (6,7,8) also meets the following Markovianity property:

$$p\left(X_{st}=x_{st} \mid X_{qr}=x_{qr}, s \neq q, t \neq r, \forall(s,t),(q,r) \epsilon V\right)=p\left(X_{st}=x_{st} \mid X_{qr}=x_{qr}, s \neq q, t \neq r,(q,r) \epsilon \eta_{sr}\right) \quad (6)$$

$$z^{g}=\frac{N^{g}}{\sum_{i=1}^{N} f\left(\frac{n_{i}^{\text{eff}}}{s_{i}}\right)\left(n_{i}^{g}\right)^{\text{eff}}} \quad (7)$$

$$\left(n_{i}^{g} \text{ eff }=\sum_{j}\left[\left(1-p^{g}\right) \delta_{ij}+p^{g} R_{ji}^{g}\right] n_{j}^{g}.\right.$$

$$f(x)=1+\left(1-e^{-\xi x}\right)$$

$$n_{j \rightarrow i}^{m,h}(t)=n_{j}^{h} \rho_{j}^{m,h}(t)\left[\left(1-p^{h}\right) \delta_{ij}+p^{h} R_{ji}^{h}\right], m \in\{A, I\} \quad (8)$$

Prior probability $P(X_t = x_t, \theta)$ is of the form and conforms to the Gibbs distribution by eqn (9).

$$P\left(X_{t}=x_{t}\right)=e^{-U\left(x_{t}, \theta\right)}=e^{\left[-\sum_{s,t}\left[V_{sc}\left(x_{st}, x_{qt}\right)+V_{tec}\left(x_{st}, x_{qr}\right)+V_{teec}\left(x_{st}, x_{er}\right)\right]\right]} \quad (9)$$

What is written for the comparable edgeless model is by eqn (10)

$$P\left(X_{t}=x_{t}\right)=e^{-U\left(x_{t}, \theta\right)}=e^{\left[-\sum_{c \epsilon C}\left[V_{sc}\left(x_{st}, x_{qt}\right)+V_{tec}\left(x_{st}, x_{qr}\right)\right]\right]} \quad (10)$$

Corresponding edgeless method is given as by eqn (11)

$$P\left(X_{t}=x_{t}\right)=e^{-U\left(x_{t}, \theta\right)}=e^{\left[-\sum_{c \epsilon C}\left[V_{sc}\left(x_{st}, x_{qt}\right)+V_{tec}\left(x_{st}, x_{qr}\right)\right]\right]} \quad (11)$$

The likelihood function $P(Y_t = y_t \mid X_t = x_t)$ can be expressed as eqn (12)

$$P\left(Y_{t}=y_{t} \mid X_{t}=x_{t}\right)=P\left(y_{t}=x_{t}+n \mid X_{t}=x_{t}, \theta\right)=P\left(N=y_{t}-x_{t} \mid X_{t}=x_{t}, \theta\right) \quad (12)$$

Thus, $P(Y_t = y_t \mid X_t = x_t)$ can be expressed as eqn (13)

$$P\left(N=y_{t}-x_{t} \mid X_{t}, \theta\right)=\frac{1}{\sqrt{(2\pi)^{f} \det[k]}} e^{-\frac{1}{2}\left(y_{t}-x_{t}\right)^{T} k^{-1}\left(y_{t}-x_{t}\right)} \quad (13)$$

$$(x)=\sum_{j=1}^{K} \pi_{j} p\left(x ; \theta_{j}\right), \; j=1, \ldots, K.$$

$$p(\mathbf{x})=\sum_{c=1}^{C} \pi_{c} f_{c}(\mathbf{x} \mid \theta) \quad (14)$$

Mixture model has a vector of parameters, $\underline{\theta}=\{\theta_1, \ldots, \theta_k, \pi_1, \ldots \pi_k\}$

Hidden variables are treated as a latent variable, or Z, in mixture models. It accepts numbers 1 through K as a discrete set that satisfies the conditions $z_k \epsilon\{0,1\}$ and $\sum_z z_k = 1$. A conditional distribution p(x | z) and a marginal distribution p(z) are how we define the joint distribution p(x, z) by eqn (15).

$$p(z, x)=p(z) p(x \mid z)$$

$$p\left(z_{k}=1\right)=\pi_{k} \quad (15)$$

A definition of probability density function of X by eqn (16)

$$p\left(x \mid \mu_{k}, \Sigma_{k}\right)=\frac{1}{\sqrt{2 \pi\left|\Sigma^{-1}\right|}} \exp \left(-\frac{1}{2}\left(x-\mu_{x}\right) \Sigma_{x}^{-1}\left(x-\mu_{x}\right)^{T}\right)$$

$$f_{c}\left(\mathbf{x} \mid \boldsymbol{\mu}_{c}, \boldsymbol{\Sigma}_{c}\right)=\frac{1}{(2 \pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}_{c}\right|^{\frac{1}{2}}} \exp \left(-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_{c}\right)^{t} \boldsymbol{\Sigma}_{c}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_{c}\right)\right) \quad (16)$$

A linear superposition of Gaussians is utilized to represent a Gaussian mixture distribution in the form by eqn (17),

$$p(x)=\sum_{k=1}^{K} \pi_{k} p\left(x \mid \mu_{k}, \Sigma_{k}\right)$$

$$\hat{\pi}_{c}=\frac{n_{c}}{n},$$

$$\hat{\mu}_{c}=\frac{1}{n_{c}} \sum_{\{f \mid y_{i}=c\}} \mathbf{x}_{i}$$

$$\hat{\boldsymbol{\Sigma}}_{c}=\frac{1}{\left(n_{c}-1\right)} \sum_{\{i \mid y,=c\}}\left(\mathbf{x}_{i}-\boldsymbol{\mu}_{c}\right)\left(\mathbf{x}_{i}-\boldsymbol{\mu}_{c}\right)^{t} \quad (17)$$

Given a certain value of z, conditional distribution of x is now a Gaussian by eqn (18):

$$p\left(x \mid z_{k}=1\right)=p\left(x \mid \mu_{k}, \Sigma_{k}\right)$$

$$p(x \mid z)=\prod_{k=1}^{K} p\left(x \mid \mu_{k}, \Sigma_{k}\right)^{z_{k}} \quad (18)$$

By adding joint distribution of all possible states of z, one can obtain the marginal distribution of x by eqn (19).

$$p(x)=\sum_{z} p(z) p(x \mid z)=\sum_{k=1}^{K} \pi_{k} p\left(x \mid \mu_{k}, \Sigma_{k}\right) \quad (19)$$

The "posterior probability" on a mixture component for a certain data vector is a significant derived quantity (20):

$$\gamma\left(z_{k}\right) \equiv p\left(z_{k}=1 \mid x\right)=\frac{p\left(z_{k}=1\right) p\left(x \mid z_{k}=1\right)}{\sum_{j=1}^{K} p\left(z_{j}=1\right) p\left(x \mid z_{j}=1\right)}=\frac{\pi_{k} p\left(x \mid \mu_{k}, \Sigma_{k}\right)}{\sum_{j=1}^{K} \pi_{j} p\left(x \mid \mu_{j}, \Sigma_{j}\right)} \quad (20)$$

Fuzzy Q- bayes gradient vector neural network:

Type-I fuzzy sets are what these fuzzy sets fall under. However, because these fuzzy sets' membership functions are sharp, they cannot be used to model many kinds of uncertainties by eqn (21).

$$A^{\prime}=\left\{(x, \mu), \mu_{A^{\prime}}(x, \mu) \mid \forall x \in U, \mu \in[0,1]\right\} \quad (21)$$

It is easy to create type-II fuzzy sets by first creating a type-I set and giving each element a lower and upper membership degree to create footprint of uncertainty (FOU), or range between lower and upper membership values. A type-II fuzzy set is what follows by eqn (22):

$$A' = \{(x, \mu_U(x), x, \mu_L(x)) \mid \mu_L(x) \leq \mu(x) \leq \mu_U(x), \mu \in [0,1]\} \quad (22)$$

where μL and μU stand for initial membership function μ(x), lower and upper membership degrees, respectively by eqn (23).

$$\mu_L(x) = [\mu(x)]^u$$
$$\mu_U(x) = [\mu(x)]^{\frac{1}{\alpha}} \quad (23)$$

where α ∈ (1,∞). Sigmoid activation function is a popular activation function in multilayer perceptrons. For binary classification issue, the sigmoid function produces continuous values in range [0, 1] that indicate likelihood of each class. Neural network may learn more complicated characteristics because sigmoid function introduces non-linearity in buried layers by eqn (24).

$$\operatorname{sig}(x) = \frac{1}{1 + e^{-x}}$$
$$\varphi_L(x) = \left[\frac{1}{1 + e^{-x}}\right]^{\alpha}$$
$$\varphi_U(x) = \left[\frac{1}{1+e^{-x}}\right]^{\frac{1}{2}} \quad (24)$$

Following definition applies to the proposed fuzzy gradient descent by eqn (25):

$$\mathbf{w} = \mathbf{w} - \text{mean}\left(|\mathbf{u}_1 - \mathbf{u}_2|^2\right) \times \eta \times \frac{d}{d\mathbf{w}} \varphi \quad (25)$$

ϕ is sort II fluffy sigmoid capability, and mean is utilized to address square distinctions between levels of participation with a solitary worth, which can be seen as a vagueness boundary. Notice that for vague hubs |u1 − u2| 2 will assess to 0, along these lines significantly affecting how loads are being refreshed. Consolidating levels of enrollment in enhancement will decide how info tests add to the learning system in view of their vagueness, with the end goal that more equivocal highlights will meaningfully affect learning, and will rather be founded on additional non-questionable elements The expense capability utilized in our work is essentially addressed as the contrast between the genuine qualities and the anticipated qualities by eqn (26).

$$\max Q(a) = \sum_{i=1}^{n} a_i - 0.5 \sum_{i,j=1}^{n} a_i, a_j, y_i, y_j K\left(x_i, x_j\right) \quad (26)$$

It has the effect of regulating the severity of the wrong classification sample's penalty and maintaining a balance between the amount of the incorrect sample and the algorithm's complexity. As shown in equation (27) the relevant discriminant function is also modified.

$$\int(x) = \operatorname{sgn} \mid \sum_{i=1}^{n} a_i^* y_i K(x_i + x) + b^* \quad (27)$$

$$q_\pi(s, a) = \mathbb{E}_\pi[G_t \mid S_t = s, A_t = a]$$

Likewise, action-value function is broken down into following parts: $q_\pi(s, a) = \mathbb{E}_\pi[R_{t+1} + \gamma q_m(S_{t+1}, A_{t+1}) \mid S_t = s, A_t = a]$

Ra s is defined as Ra s = E[Rt+1|St = s, At = a] to streamline notations. Additionally, we may observe the connection between v(s) and q(s, a) by eqn (28):

$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a \mid s) q_\pi(s, a)$$
$$q_\pi(s, a) = \mathcal{R}_s^a + \gamma \sum_{n' \in \mathcal{S}} \mathcal{P}_{ss'}^a v_\pi(s')$$
$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a \mid s) \left(\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v_*(s')\right) \quad (28)$$

By maximising q(s, a) over all actions, we can use the theorem to determine an ideal course of action right away. $\pi_*(a \mid s) = \begin{cases} 1 & \text{if } a = \arg\max_{a \in A} q_*(s, a), \\ 0 & \text{otherwise.} \end{cases}$

According to Bayes method: $P(C \mid X) = \frac{P(X|C)P(C)}{P(X)}$

The goal of the Naive Bayes classifier is to assign the text to the category v that best fits it based on the text vector $X(x_1, x_2 \ldots x_n) \cdot X(x_1, x_2 \ldots x_n)$is the text's feature vector, and $C_1, C_2 \ldots C_j$ are the categories that have been established. The calculation is therefore concerned with the probability $(P_1^-, P_2 \ldots P_n)$ when $X(x_1, x_2 \ldots x_n)$ belongs to $C_1, C_2 \ldots C_j$, with P_j being the probability when $X(x_1, x_2 \ldots x_n)$ belongs to C_j. Then max $(P_1, P_2 \ldots P_n)$ is the desired outcome then. Naive Bayes method allows for the following formula to be obtained by eqn (29):

$$P\left(C_j \mid x_1, x_2, \ldots, x_n\right) = P\left(x_1, x_2, \ldots, x_n \mid C_j\right) P\left(C_j\right) \quad (29)$$

P(Cj) represents prior probability that text belongs to Cj in this formula, while P(x1,x2... xn|Cj) represents posterior probability that Cj contains text vector (x1,x2... xn) when text to be categorised belongs to Cj. Therefore, the greatest value of following formula is equivalent to max (P1,P2,...,Pn) by eqn (30):

$$\operatorname*{argmax}_{C_j \in C} P\left(x_1, x_2, \ldots, x_n \mid C_j\right) P\left(C_j\right) \quad (30)$$

The qualities (x1,x2,...,xn) are assumed by Bayes to be independent of one another. Product of probabilities for every attribute is then joint probability. Following is the final classification method by eqn (31):

$$\underset{C_j \in C}{\operatorname{argmax}} P(C_j) \prod_{i=1}^{n} P(x_i \mid C_j) \quad (31)$$

As a rule, when given a Multilingual text, we don't know which class it has a place with, which requires the text to get a similar earlier likelihood for every class as indicated by the standard of reasonableness. Because there are different types of texts in the training sets, it is unfair as well as unreasonable to consider prior probability to be different. Consequently, it is sensible to eliminate the computation of earlier likelihood and equivalents to relegate a similar earlier likelihood. The following classification function can then be obtained by eqn (32):

$$\arg\ \max \prod^{n} P(x_i \mid C_j) \quad (32)$$

Since the most extreme likelihood is the thing is required, eliminating the estimation of earlier likelihood won't influence the last grouping result, however extraordinarily accelerate the computation. The study used a double-valued variable to represent the posterior probability. Some of the time the text to be ordered is extremely lengthy, making component of text vector significantly more as well as back likelihood tiny. Also, after every one of the characteristics' probabilities are increased, the outcome might be zero, so a mistake engendering shows up. To tackle this issue, we extend the back likelihood of each element property to a certain numerous. This won't influence the examination results, since what makes a difference in the end is the correlation of likelihood among classes and it is more helpful on the off chance that the likelihood is duplicated. From the beginning, we amplified multiple times; in any case, in analysis, we found that occasionally back likelihood will past extent of variable of twofold, which will extraordinarily influence the trial results. It is possible to create k classifiers with k parts using the SVM technique for parallel order, and m-th bit type of SVM limit between m-th class as well as leftover (k 1) classes can be created as follows by eqn (33).

$$D_m(\mathbf{x}) = \sum_{i \in SV_m} \alpha_i^{(m)} y_i^{(m)} K_m(\mathbf{x}_i, \mathbf{x}) + b^{(m)} \quad (33)$$

One is the Gaussian Radial Basis Function kernel, which is a radial kernel $K(\mathbf{x}, \mathbf{x}') = \tilde{f}(-\|\mathbf{x} - \mathbf{x}'\|^2/2)$ by eqn (34).

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\|\mathbf{x} - \mathbf{x}'\|^2/2\sigma^2\right) \quad (34)$$

For a variety of application fields, such as text mining and bioinformatics, for instance, special kernel functions have been developed. The interpretation of a normalised kernel function for a pair of points in the feature space is as follows by eqn (35):

$$\cos\ \theta_{\varphi(x),\varphi(z)} = \frac{\varphi(x)^T \varphi(z)}{\|\ \varphi(x)\ \|_2 \|\ \varphi(z)\ \|_2} = \frac{K(x,z)}{\sqrt{K(x,x)}\sqrt{K(z,z)}}$$

$$\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2} w^T w + c \sum_{i=1}^{N} (\xi_i + \xi_i^*)$$

$$\text{subject to } y_i - w^T \varphi(x_i) - b \le \epsilon + \xi_i,\ i = 1, \ldots, N$$

$$^T \varphi(x_i) + b - y_i \le \epsilon + \xi_i^*,\ i = 1, \ldots, N$$

$$\xi_i, \xi_i^* \ge 0,\ i = 1, \ldots, N.$$

$$\text{(D) } \max_{\alpha,\alpha^*} -\frac{1}{2} \sum_{i,j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) \begin{matrix} -\epsilon \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) + \\ \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0 \end{matrix}$$

$$\text{subject to} \quad \begin{matrix} \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, c], i = \end{matrix} \quad (35)$$

It has box restrictions and is a quadratic programming problem.

4. Performance analysis:

On a computer with a Core i7 processor and 16 GB of RAM, we conducted our research. Spyder, part of the Anaconda software suite, was used to assess our suggested prediction models. Additionally, we split our dataset into training and testing using stratified 5-fold cross-validation and random hold-out as two strategies for model validation.

Data Description: The informational index utilized in this paper is taken from a student action tracker device experience Programming interface (xAPI). Learning and training architecture (TLA) provided by xAPI makes it easier to track a student's progress in learning and monitor their activities, like writing an article, watching a video, or reading an article, among other things. xAPI gives learning stage facilitator to recognize student, activities and all connected components that could end up being useful to in tending to the learning practice. The dataset comprises of 480 understudy records and 16 credits. There are three main categories for these characteristics: 1) Segment related ascribes like orientation and identity. (2) Scholarly related credits like instructive stage, grade Level and segment. (3) Conduct related ascribes like hand ascending in classes, opening assets, doing overviews with cohorts, and school fulfillment. The dataset is gathered through 2 intellectual. In Europe, for example, two organizations financed by European Commission (European Master Organization on Financial matters of Schooling - EENEE, and Organization of Specialists on Friendly Parts of Instruction and Preparing - NESET) are answerable for composing Reports for Commission, that then proceed to impact its true correspondences. Pisa in Focus is a concise note distributed by OECD with policy recommendations based on data analyses of its Program for International Student Assessment (PISA).

Table-1 Analysis based on various students' academic dataset

Dataset	Techniques	Training Accuracy	Average Precision	Recall	MSE
EENEE	Naïve bayes	82	70	56	45
	Decision tree	85	75	64	43
	SMG_FQBGVNN	87	84	67	40
NESET	Naïve bayes	89	72	63	49
	Decision tree	94	76	68	45
	SMG_FQBGVNN	97	83	74	39
PISA	Naïve bayes	90	89	72	55
	Decision tree	93	93	75	50
	SMG_FQBGVNN	99	94	79	49

Table-1 shows analysis based on various students academic dataset. Here dataset analysed are EENEE, NESET, PISA dataset in terms of training Accuracy, average Precision, Recall, MSE.

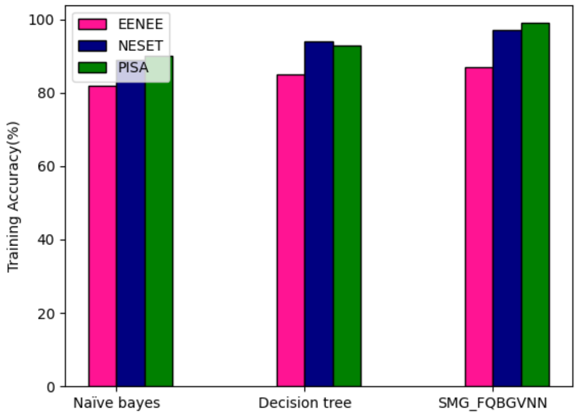

Figure-2 Comparison of training accuracy

Figure 2 shows analysis of training accuracy. Here proposed technique attained 87%, existing naïve bayes attained 82%, decision tree attained 85% for EENEE dataset; for NESET dataset the proposed technique attained 97%, existing naïve bayes attained 89%, decision tree attained 94%; the proposed technique attained 99%, existing naïve bayes attained 90%, decision tree attained 99% for EENEE dataset.

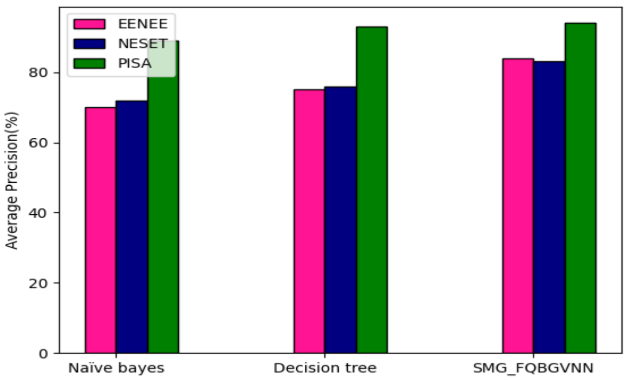

Figure-3 Comparison of average precision

From above figure 3 shows comparative analysis of average precision for various students academic performance dataset. the proposed technique attained 84%, existing naïve bayes attained 70%, decision tree attained 75% for EENEE dataset; for NESET dataset the proposed technique attained 83%, existing naïve bayes attained 72%, decision tree attained 76%; the proposed technique attained 94%, existing naïve bayes attained 89%, decision tree attained 93% for EENEE dataset.

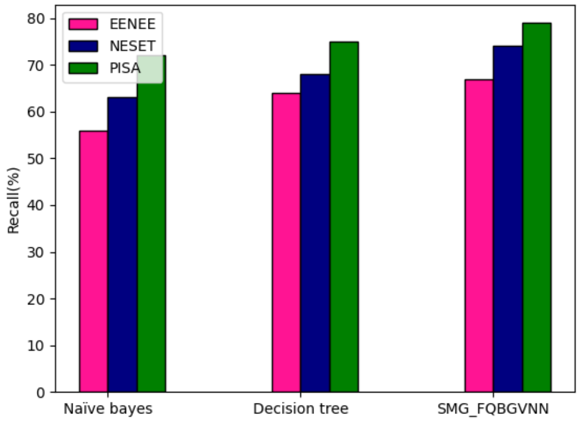

Figure-4 Comparison of Recall

Figure 4 shows analysis of Recall. Here proposed technique attained 67%, existing naïve bayes attained 56%, decision tree attained 64% for EENEE dataset; for NESET dataset the proposed technique attained 74%, existing naïve bayes attained 63%, decision tree attained 68%; the proposed technique attained 79%, existing naïve bayes attained 72%, decision tree attained 75% for EENEE dataset.

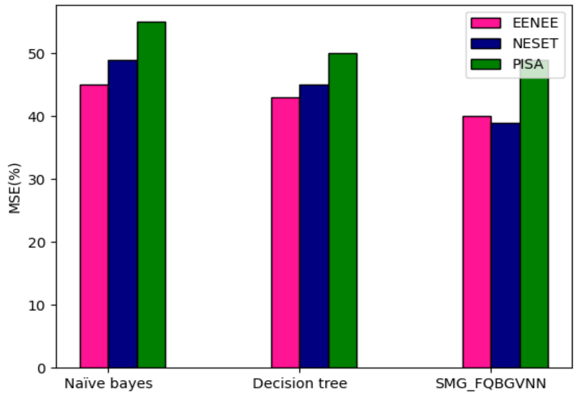

Figure-5 Comparison of MSE

Figure 5 shows analysis of MSE for various students academic performance dataset. Proposed technique attained 40%, existing naïve bayes attained 45%, decision tree attained 43% for EENEE dataset; for NESET dataset the proposed technique attained 39%, existing naïve bayes attained 49%, decision tree attained 45%; the proposed technique attained 49%, existing naïve bayes attained 55%, decision tree attained 50% for EENEE dataset.

We next present the outcomes from our disconnected examinations for the proposed models, and afterward talk about the compromises and plan choices to pick a model for online trial and error. We at long last present the consequences of our online A/B trial of the picked model on LinkedIn Spotter item, which depends on unaided embeddings. More than 60,000 sentences were collected all together, including both sentences mentioned in the "Required Skills" section and sentences mentioned in other vacancies sections. We prepared a twofold classifier model involving FastText library in Python for our grouping task. As part of its training, classifier uses word n-grams as well as learns embeddings. The got dataset was parted as 80% for preparing and 20% for assessment.

Reproduction system: (1) For every conceivable number of characteristic qualities ($2 \leq |A| \leq 10$): (a) Create a set P of 100K irregular unmitigated likelihood circulations of size |A| each. Every likelihood dissemination $Pj \in P$ is created by picking |A| i.i.d. tests from uniform dissemination north of (0, 1) and normalizing aggregate to approach 1. A possible desired distribution of attribute values over set A is represented by each Pj. b). We repeat this step ten times, producing 1,000,000 distinct ranking tasks for each selection of |A|. ii) Run each proposed decency mindful positioning calculation to get a decency mindful re-positioned rundown of size 100, with the ideal dissemination Pj and the produced irregular up-and-comer records for each trait esteem as data sources. For each positioning undertaking produced by above system, we process proposed inclination measures like InfeasibleIndex (Eq. 8), MinSkew (Eq. 2), and NDKL (Eq. 4), as well as Standardized Limited Total Gain1 (NDCG) as a proportion of "positioning utility" where we treat scores of up-and-comers as their importance.

An examination of accomplished precision versus gauge exactness to choose if the information was adequate. On the off chance that the pattern information was lower than the accomplished exactness via preparing the model, then, at that point, how much information utilized was considered to be adequate. Total number of TRUE data points (TRUE = giving a value of "1" = cleared course) was divided by number of the larger set to determine baseline accuracy. A reason by analogy approach, which means estimating the amount of data required by examining the outcomes of comparable previous applied machine learning studies. A strategy that is utilized to decide the exactness of a ML model is to gauge the precision of the prepared model against the pattern precision of a similar model. On the off chance that the prepared precision is higher, the model is considered to be sufficient for the information utilized. To work on the precision of determining, a brain network with a three-layer design, comprising of an information, stowed away and yield layer, was utilized. A single neuron with the logistic softmax activation function made up the output neural layer, which indicated whether the student had joined the cluster of successful or unsuccessful students. The quantity of neurons and enactment elements of the information and secret layers were chosen consequently founded on the keras-tuner library. According to the study, all aspects of the student must be taken into account in order to accurately predict their final academic performance. low connection with the objective notwithstanding, during the forecast it contribute altogether. In conclusion, the proposed model outperformed benchmark studies with comparable data sets, meeting the study's objectives. There is still room for improvement despite proposed model's superior performance. The model requires additional investigation on additional data sets, among other limitations of the proposed study. The data set is also affected by the imbalance. In this manner, the model should be additionally approved utilizing a huge and adjusted informational collection.

5. Conclusion:

This research propose novel technique in students performance utilizing machine learning based on feature extraction and classification using spatio markov Gaussian model (SMG) and fuzzy Q- bayes gradient vector neural network (FQBGVNN). The model is a tool for anticipating student's grades in advance. Review demonstrated meaning of DL method in EDM. Destroyed was applied to decrease gamble of method overfitting. Proposed method beat when contrasted and benchmark concentrates on utilizing similar informational index. According to the findings of the study, each and every one of the features makes a significant contribution to predicting academic performance. Moreover, the profound learning models has enormously upgraded the forecast execution. Students' academic performance was therefore predicted using a variety of predictors, algorithms, and methods. The findings demonstrate that students' academic performance can be predicted utilizing ML algorithms. All the more critically, the expectation was made exclusively with the boundaries of midterm grade, staff and division. Educating staff can benefit from the consequences of this exploration in the early acknowledgment of understudies who have beneath or better than expected scholarly inspiration. Future study can be conducted by considering more parameters as input variables and adding more ML techniques to method process. Moreover, it is important to outfit efectiveness of DM techniques to examine understudies' learning ways of behaving, address their concerns, streamline instructive climate, and empower information driven navigation.

Statement and Declarations

Ethical Approval: This article does not contain any studies with animals performed by any of the authors.

Competing Interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Fund information: No Funding

Availability of data and Materials: All the data’s available in the manuscript

Reference:

[1] Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11.

[2] Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. *Decision Analytics Journal*, 100204.

[3] Nti, I. K., Akyeramfo-Sam, S., Bediako-Kyeremeh, B., & Agyemang, S. (2022). Prediction of social media effects on students’ academic performance using Machine Learning Algorithms (MLAs). *Journal of Computers in Education*, *9*(2), 195-223.

[4] Zhang, W., Wang, Y., & Wang, S. (2022). Predicting academic performance using tree-based machine learning models: A case study of bachelor students in an engineering department in China. *Education and Information Technologies*, *27*(9), 13051-13066.

[5] Fahd, K., Venkatraman, S., Miah, S. J., & Ahmed, K. (2022). Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature. *Education and Information Technologies*, 1-33.

[6] Ouatik, F., Erritali, M., Ouatik, F., & Jourhmane, M. (2022). Predicting student success using big data and machine learning algorithms. *International Journal of Emerging Technologies in Learning (Online)*, *17*(12), 236.

[7] Chen, Y., & Zhai, L. (2023). A comparative study on student performance prediction using machine learning. *Education and Information Technologies*, 1-19.

[8] Al-Alawi, L., Al Shaqsi, J., Tarhini, A., & Al-Busaidi, A. S. (2023). Using machine learning to predict factors affecting academic performance: the case of college students on academic probation. *Education and Information Technologies*, 1-26.

[9] Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2022). Predicting students' academic performance using machine learning techniques: a literature review. *International Journal of Business Intelligence and Data Mining*, *20*(4), 456-479.

[10] Hussain, S., & Khan, M. Q. (2023). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. *Annals of data science*, *10*(3), 637-655.

[11] Verma, U., Garg, C., Bhushan, M., Samant, P., Kumar, A., & Negi, A. (2022, March). Prediction of students' academic performance using Machine Learning Techniques. In *2022 International Mobile and Embedded Technology Conference (MECON)* (pp. 151-156). IEEE.

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

[13] Chen, S., & Ding, Y. (2023). A Machine Learning Approach to Predicting Academic Performance in Pennsylvania's Schools. *Social Sciences*, *12*(3), 118.

[14] Xu, K., & Sun, Z. (2023). Predicting academic performance associated with physical fitness of primary school students using machine learning methods. *Complementary Therapies in Clinical Practice*, 101736.

[15] Hasib, K. M., Rahman, F., Hasnat, R., & Alam, M. G. R. (2022, January). A machine learning and explainable ai approach for predicting secondary school student performance. In *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0399-0405). IEEE.

[16] Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. *Materials today: proceedings*, *80*, 3782-3785.

[17] Atlam, E. S., Ewis, A., Abd El-Raouf, M. M., Ghoneim, O., & Gad, I. (2022). A new approach in identifying the psychological impact of COVID-19 on university student's academic performance. *Alexandria Engineering Journal*, *61*(7), 5223-5233.

[18] Olabanjo, O. A., Wusu, A. S., & Manuel, M. (2022). A machine learning prediction of academic performance of secondary school students using radial basis function neural network. *Trends in Neuroscience and Education*, 100190.

[19] Sai Charan, N., Ali Hussain, M., Vineela, P., Vamsi Adi Tilak, M., & Chandu Siva Shankar, T. (2022, May). Predictive Student Performance Analysis Using Machine Learning and Student Assistance System. In *ICCCE 2021: Proceedings of the 4th International Conference on Communications and Cyber Physical Engineering* (pp. 1105-1113). Singapore: Springer Nature Singapore.

[20] Suleiman, R., & Anane, R. (2022, May). Institutional data analysis and machine learning prediction of student performance. In *2022 IEEE 25th international conference on computer supported cooperative work in design (CSCWD)* (pp. 1480-1485). IEEE.