Taylor & Francis
Taylor & Francis Group

Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/taut20

Effective deep learning based grade prediction system using gated recurrent unit (GRU) with feature optimization using analysis of variance (ANOVA)

S. Lakshmi & C. P. Maheswaran

To cite this article: S. Lakshmi & C. P. Maheswaran (2024) Effective deep learning based grade prediction system using gated recurrent unit (GRU) with feature optimization using analysis of variance (ANOVA), Automatika, 65:2, 425-440, DOI: 10.1080/00051144.2023.2296790

To link to this article: https://doi.org/10.1080/00051144.2023.2296790

AUTOMATIKA
2024, VOL. 65, NO. 2, 425–440
https://doi.org/10.1080/00051144.2023.2296790

Effective deep learning based grade prediction system using gated recurrent unit (GRU) with feature optimization using analysis of variance (ANOVA)

S. Lakshmi[a] and C. P. Maheswaran[b]

[a]Department of Computer Science and Engineering, Noorul Islam Centre for Higher Education, Kumaracoil, India; [b]Department of Artificial Intelligence and Data Science, Sri Krishna College of Technology, Coimbatore, India

ABSTRACT
The prediction of final semester grades is a crucial undertaking in education, offering insights into student performance and enabling timely interventions to support their academic journey. This paper employs a deep learning approach, specifically gated recurrent unit (GRU), in conjunction with feature optimization using analysis of variance (ANOVA), to forecast final semester grades. The predictive model is trained and evaluated on a handcrafted grade prediction dataset, which contains the academic performance of the students during Plus 2 and from Semester 1 to Semester 4 of a group of Computer science and Engineering majors in Kerala. By processing historical academic records and contextual information, the GRU model learns to predict future performance accurately. To enhance the model's efficacy and interpretability, ANOVA is applied to optimize the feature selection process. This statistical technique identifies the most influential factors contributing to final grades, refining the model's predictive power while reducing dimensionality. The experimental results showcase the model's effectiveness in predicting final semester grades, demonstrating superior accuracy and performance compared to grade prediction using CNN with Bayesian optimization and LSTM with L1-Norm optimization.

ARTICLE HISTORY
Received 22 September 2023
Accepted 10 December 2023

KEYWORDS
Student performance prediction; artificial intelligence; deep learning; recurrent neural network; gated recurrent unit; feature optimization; ANOVA

1. Introduction

In recent times, the exponential advancement of technology has led to an unprecedented surge in data across various sectors. The demand for uncovering fresh and valuable insights from these vast data pools has also surged. With the emergence of data mining, diverse mining methods have been deployed across various industries, including education, finance, retail, bioinformatics and telecommunications. These techniques aim to extract valuable information to meet the evolving needs of these sectors [1]. In the era of vast data repositories comprising files, databases and other storage systems, there is a crucial need to cultivate robust methods for the analysis and comprehension of this data. Such methods are essential for extracting valuable insights that can inform decision-making processes, ultimately yielding meaningful and valuable information [2]. Graduate education holds a pivotal role within the national education system, serving as a cornerstone for fostering competitiveness and driving innovation. A primary focus of higher education is the ongoing assessment and improvement of educational quality. Education stands as a cornerstone for a nation's advancement, facilitating the spread and enrichment of key ideas. Educational institutions and the process of learning wield tremendous influence in molding individuals, encompassing not only their intellect but also their overall personality [3]. Educational institutions play a significant role in fostering a strong cultural and linguistic identity while also serving as crucial hubs for skill development and academic excellence. In today's interconnected world, a robust higher education system is vital for a nation's economic advancement. A well-functioning higher education system not only aligns with but also enhances the pathways to economic and social development, ultimately paving the way for a brighter future for the nation's populace.

Over the past decade, education in India has experienced significant expansion, largely driven by the government's support and encouragement in the establishment of new educational institutions [4]. As a result, educational institutions have been compelled to reevaluate their scope and objectives to ensure their long-term sustainability. Virtually all educational institutions are overseen by regulatory bodies that play a vital role in establishing guidelines for infrastructure design, the acquisition of state-of-the-art software and equipment, the selection of qualified educators and the implementation of strategies to motivate and enhance their performance. However, people find themselves in an intensely competitive era of research and development, where knowledge holds significant power.

CONTACT S. Lakshmi lakshmis.cse@gmx.com Department of Computer Science and Engineering, Noorul Islam Centre for Higher Education, Kumaracoil, Tamil Nadu, India

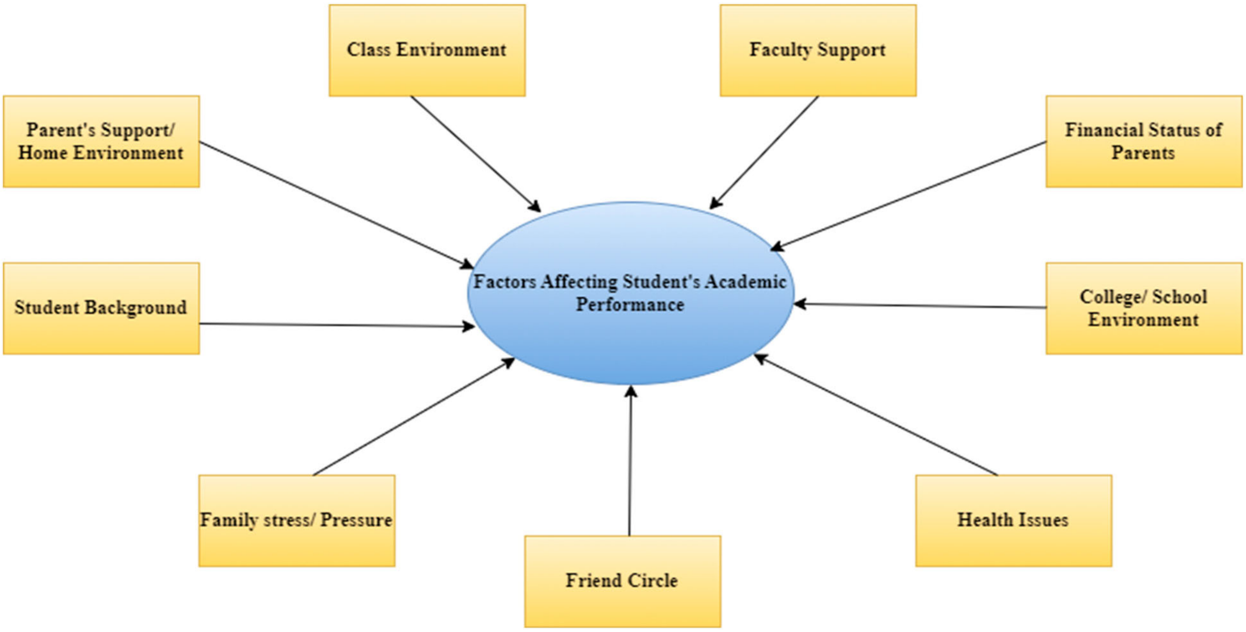

Figure 1. Factors affecting student's academic performance.

Profound insights into the performance of both educators and students have the potential to elevate the quality of education, thereby aiding institutions in bolstering student enrolment, increasing retention rates and enhancing educator effectiveness, ultimately elevating the overall performance of the institution. The rising prevalence of rote learning and the growing demand for well-rounded individuals have prompted administrators and researchers in educational institutions to explore alternative teaching methods and the delivery of high-quality education. The provision of such education inherently enhances students' effectiveness and capabilities.

Within the realm of higher education, a student's academic performance is a crucial gauge of their advancement and future prospects. Figure 1 illustrates the various elements that impact a student's academic achievement.

A persistent challenge within the realm of higher education revolves around ensuring that students remain enrolled and ultimately achieve successful graduation [5]. The National Research Council's report [6] has highlighted a pressing need for the development of innovative methods that can assist higher-education institutions in retaining students, ensuring their timely graduation and preparing them effectively for the workforce in their chosen fields of study. A crucial aspect of achieving these objectives is the ability to forecast student grades in upcoming enrolment periods, which can offer valuable insights for students, academic advisors and educators. Predicting grades for the next semester or even multiple semesters ahead can aid students in making informed decisions about their majors, optimizing their course schedules by balancing challenging and easier courses, and signalling to advisors and educators when students may require additional support. However, this task is becoming increasingly complex due to the growing volume of data generated by rising student enrolments and the evolving characteristics of the student population.

The prediction of final semester grades for students is an essential aspect of academic assessment, and its significance extends far beyond the confines of the classroom. In particular, the accurate anticipation of a student's performance in their last semester carries substantial implications for their career advancement and long-term professional growth. As students approach the culmination of their academic journey, the grades they attain serve as a critical tool, influencing various facets of their professional trajectory. By analyzing various factors, such as previous academic records, attendance and assignment scores, the grade prediction system aims to provide educators and administrators with valuable insights that can inform instructional strategies, identify at-risk students and enable timely interventions to ensure each student's academic success. Traditional approaches for grade prediction in education have historically relied on simple metrics such as past academic performance, standardized test scores and teacher evaluations. These methods often lack the nuance and predictive power offered by more modern data-driven techniques. They tend to overlook individual student characteristics, engagement levels and the broader context of the learning environment. While these traditional approaches can provide a general sense of a student's academic potential, they often fall short in offering personalized insights, early identification of struggling students, or targeted interventions to improve educational outcomes. In this paper, a GRU-based multiclassifier model was introduced for final semester grade prediction based on plus two marks and student's academic data up to the fourth semester.

The main contributions of the proposed work includes:

- Final semester grade prediction of students with the help of GRU model.
- Feature optimization was performed with the help of ANOVA.
- Performance comparison of the proposed model with other deep learning approaches.

The rest of the paper is structured in the following way: The literature review and research gaps are provided in section two. The detailed methodology is highlighted in section three. Section four discusses the comprehensive findings of the suggested approach as well as the performance comparison. Section five presents conclusion of the work.

2. Background

2.1. Literature review

D. Sreenivasulu et al. [7] developed a Convolutional Neural Network-based system for predicting students' grades. The primary objective of this work was to pinpoint the key factors that impact students' academic performance and determine the most effective approach for predicting their grades. The simulation results demonstrated that the proposed CNN model achieved an impressive accuracy rate of 98.35%.

Siti Dianah Abdul Bujang et al. [8] introduced a multiclass prediction model that utilizes six different predictive models. This model aims to forecast the final grades of students based on their previous performance in the final examination result of the first-semester course. To assess the accuracy of student grade prediction, the combination of oversampling SMOTE with various feature selection methods are explored. The results revealed that the suggested multiclass prediction model outperformed the use of oversampling SMOTE or feature selection alone under specific parameter configurations.

Yupei Zhang et al. [9] introduced a novel matrix factorization technique known as "graph regularized robust matrix factorization" (GRMF). This approach builds upon recent advancements in robust matrix factorization and incorporates two side graphs derived from student and course data into the objective function of robust low-rank matrix factorization. The evaluation of this method using relevant matrices showcased its superior performance compared to other methods. Specifically, GRMF achieved the highest accuracy of approximately 65.4% in predicting student grades.

Zoe Kanetaki et al. [10] introduced a generalized linear autoregressive model to assess and measure the primary factors influencing the academic performance of mechanical engineering students. The primary function of the proposed approach is to forecast student grades in scenarios involving online learning within hybrid educational settings. The enhanced iteration of the GLAR model demonstrates an impressive capability to predict student grades with an accuracy level of approximately ± 1, achieving a success rate of 63.70%. This marks a substantial improvement of 28.08% in accuracy compared to previously established models.

Yupei Zhang et al. [11] introduced a student performance prediction system that utilizes Sparse Attention Convolutional Neural Networks (SACNN). SACNN is designed as a sequential network, consisting of sparse attention layers, convolutional neural layers and fully connected layers. This study involved an evaluation on the NPU-CS dataset, with the simulation results demonstrating that the proposed SACNN model achieved impressive performance metrics. Specifically, it achieved an 81% F1-score and an 87% AUC for grade prediction, while it also achieving an 85% accuracy rate for student failure. These results clearly surpass the performance of existing prediction systems in this domain.

Sri Udaya Damuluri et al. [12] developed a grade prediction system employing Support Vector Machine (SVM) technology. This study utilizes a comprehensive dataset containing various student attributes. The primary objective was to forecast students' final grades and pinpoint those who might be at risk of failing by leveraging SVM in conjunction with the amassed attributes. The outcomes of this study indicated that the suggested approach achieved a commendable accuracy rate, accurately predicting the grades of approximately 70% of the students.

Dinh Thi Ha et al. [13] conducted an empirical research study, which aimed to predict student performance using machine learning methods. The dataset was collected by amalgamating information obtained through surveys of graduate students spanning three different academic years and data from the university's student management system. The simulation results demonstrated an impressive accuracy rate, with all students correctly classified. Notably, the use of MLP and Naive Bayes algorithms yielded higher true positive rates and recall values across most of the classes.

Lonia Masangu et al. [14] developed a student performance prediction model employing data mining methods. They employed a range of machine learning models including Support Vector Machine, Decision Tree, Perceptron Classifier, Logistic Regression and Random Forest Classifier. The simulation output revealed that the Support Vector Machine algorithm emerged as the most suitable choice for forecasting student academic performance, with an accuracy rate of 70.8%, surpassing other algorithms.

Efrem Yohannes Obsie and Seid Ahmed Adem [15] developed an application aimed at helping higher education institutions anticipate students' academic performance well before graduation, with the goal of reducing dropout rates. The student performance is analysed by analyzing their Cumulative Grade Point Average (CGPA) at the end of the eighth semester. Various data mining techniques are employed which includes Neural Network (NN), Support Vector Regression (SVR) and Linear Regression (LR) for predictive purposes. The simulation results demonstrated the effectiveness of these data mining methods in modelling and forecasting students' final CGPA in higher education settings, potentially serving as a valuable predictive tool.

Amal Alhassan et al. [16] conducted research into how assessment grades and online activity data within a Learning Management System (LMS) can influence students' academic performance. Five different classification algorithms are utilized including decision tree, random forest, sequential minimal optimization, multilayer perceptron and logistic regression. The findings indicated that random forest emerged as the top-performing classifier in predicting students' performance, achieving the highest accuracy scores for both the base model and sub-model, with decision tree coming in as the second-best performer.

Nalindren Naicker et al. [17] introduced a linear support vector machine as a means to forecast student achievement in the context of school-based education. The experiments are conducted to assess its effectiveness involving feature selection, using a dataset comprising 1000 student records that included both alphanumeric data. In comparison to 10 other categorical machine learning algorithms, the linear support vector machines consistently demonstrated superior predictive performance for student performance. The simulation results unveiled that certain factors, such as race and gender, played significant roles in mathematics performance.

Raghad Alshabandar et al. [18], introduced a predictive model for assessing student performance in online courses by leveraging machine learning algorithms. This study presents the development of two distinct predictive models aimed at identifying the key factors impacting students' learning outcomes in massive open online courses (MOOCs). The findings of this investigation demonstrate that both models yield practical and precise results.

Leena H. Alamri [19] introduced a model for predicting student performance aimed at improving the academic outcomes of educational institutions. This study utilized classification methods, specifically Support Vector Machines (SVM) and Random Forest (RF) algorithms. Both binary classification and regression techniques were employed with both SVM and RF. The experimental findings for both SVM and RF, applied to the two datasets, demonstrated remarkable results. In binary classification, the model achieved an impressive accuracy rate of 93%, while in regression, the lowest Root Mean Square Error (RMSE) recorded was 1.13, particularly in the case of RF.

Sana Bhutto et al. [20] introduced a predictive model for assessing student academic performance. This model employed supervised machine learning algorithms such as support vector machines and logistic regression. The research findings were backed by a series of experiments utilizing various technologies. The outcomes demonstrated that the sequential minimal optimization algorithm outperformed logistic regression, achieving significantly higher accuracy. Moreover, the results obtained from employing gain ratio feature selection in conjunction with support vector machines indicated that student satisfaction, interaction with the educational system, and punctuality in the classroom were identified as the three primary categories of factors influencing academic grades.

Sachin Rai et al. [21] designed a student performance prediction model leveraging machine learning techniques. This work specifically employed the Random Forest (RF) algorithm to forecast student marks. The results revealed that the RF classifier outperformed other classification methods, such as the Support Vector Machine (SVM), in terms of accuracy. The suggested approach holds potential for assisting both teachers and students in improving the overall quality of the learning experience.

Table 1 shows the comparison of various algorithms used for grade prediction of students. The limitations of different algorithms are also included in Table 1.

2.2. Research gap

Predicting the grades of students is a complex task influenced by various factors. Several algorithms and models can be employed for this purpose, but they come with their own limitations. Linear regression assumes a linear relationship between independent and dependent variables. If the relationship is non-linear, the model may not perform well. Linear regression is sensitive to outliers, which can significantly impact the predicted grades. Decision trees are prone to overfitting, especially when they are deep and complex. This can lead to poor generalization on new data. Small changes in the data can result in different tree structures, leading to instability in predictions. Decision trees may not handle continuous variables well without preprocessing. Random forests consist of multiple decision trees, which can make them computationally expensive and harder to interpret. SVM performance depends on the choice of the kernel function, and selecting the wrong kernel can result in poor predictions. Deep learning models, including neural networks, require large amounts of data for training. Limited data can lead to overfitting. The choice of the number of neighbours in KNN can impact the results.

Table 1. Performance comparison of various algorithms used for grade prediction of students.

Ref. no.	Algorithm	Accuracy (%)	Limitations
[22]	Wrapper based Decision Tree	94.8	• Limited use of attributes.
[23]	Deep Neural Network	89	• Feature selection is not considered.
[24]	Naive Bayes	85.6	• Small dataset and limited to one classifier. • Not generalizable.
[25]	CONV-LSTM	90	• Lacking generalization of the model.
[26]	Random Forest	75.2	• Not efficient for high dimensional imbalanced dataset.
[27]	SVM, NB, MLP, KNN, LR	97.6	• Not use an imbalanced method to address imbalanced data problems.
[28]	RF feature selection, MLP	62	• Limited sample size.
[29]	SMOTE + GWO + RF	98.8	• Not resolving data sparsity issues.
[30]	SMOTE + NB with boosting	96.1	• Feature selection is not considered.

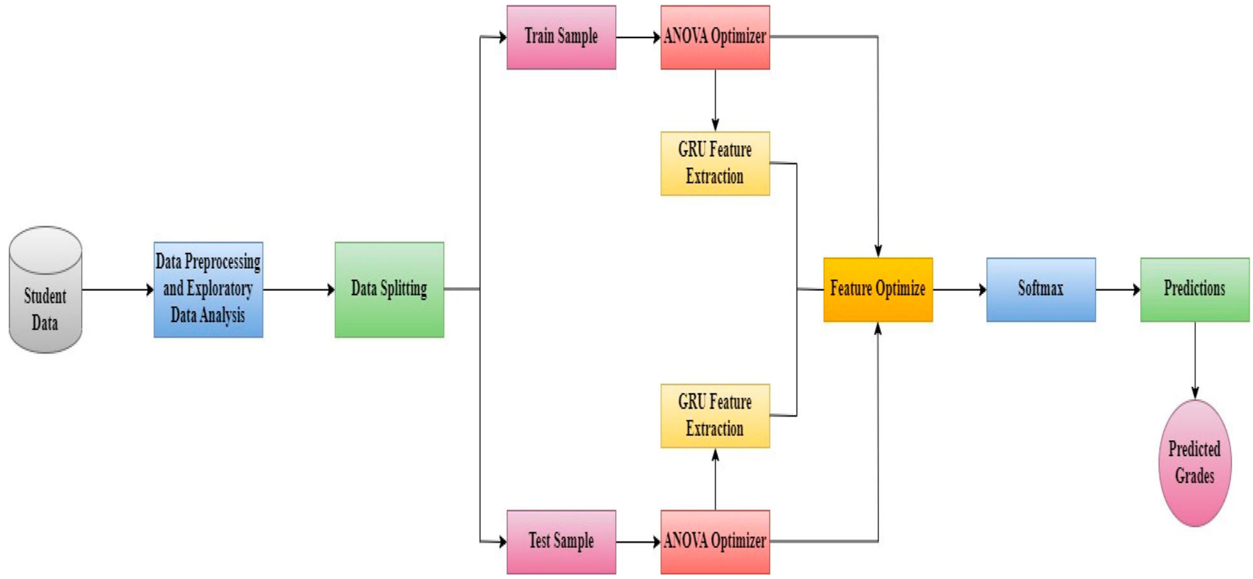

Figure 2. Block diagram of the proposed methodology.

Selecting an inappropriate k value can lead to inaccurate predictions. Addressing these limitations often requires a thoughtful combination of algorithms, feature engineering, and domain knowledge to build effective grade prediction models. So, an effective grade prediction system was presented in this paper.

3. Proposed methodology of grade prediction system

A GRU model with feature optimization using ANOVA is developed and analysed for grade prediction of students. The detailed block diagram of the proposed work is shown in Figure 2. The initial step of the work involves the dataset collection. It is followed by data preprocessing techniques. The features are optimized using the ANOVA optimizer. The optimized features are then fed into the GRU model for final prediction. Finally, the performance of the proposed prediction model is evaluated and compared with other grade prediction models.

3.1. Dataset collection

The proposed system utilizes a handcrafted dataset that tailored to the specific requirements of the system. The dataset is obtained from the student records of various engineering colleges in South India. The dataset consists of 12,081 Rows and 21 Columns. 10 different grades are provided as ground truth to train the model. The various attributes in the dataset are shown in Figure 3.

In this paper, the multiclass classifier is developed to predict the grade based on 20 inputs. The sample dataset is shown in Figure 4.

The columns indicate the academic performance of the students during Plus 2 and from Semester 1 to Semester 4. The final grade is provided as the ground truth for training and testing the GRU model. The grades can be predicted according to the CGPA range. It is tabulated in Table 2.

Table 2. Grades and CGPA ranges.

Grade	CGPA Range
S	9–10
A+	8.5–9
A	8–8.5
B+	7.5–8
B	7–7.5
C+	6.5–7
C	6–6.5
D	5.5–6
P	5–5.5
F	Less than 5 (fail)

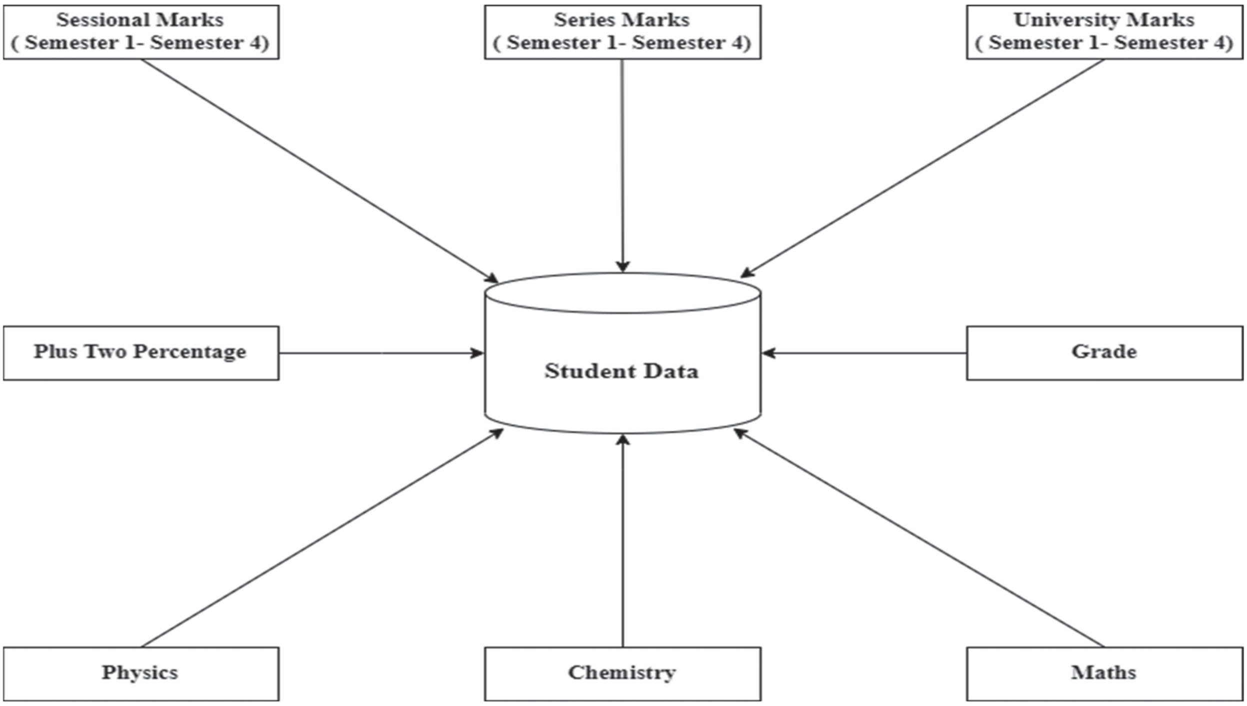

Figure 3. Various attributes in grade prediction dataset.

	Plus 2 Percentage	Physics	Chemistry	Maths	S1_Series 1	S1_Series 2	S1_Sessional	S1_University	S2_Series 1	S2_Series 2	...	S2_University	S3_Series 1	S3_Series 2	S3_Sessional	S3_University	S4_Series 1	S4_Series 2	S4_Sessional	S4_University	Grade
0	96.600000	93	97	95	222.0	189.0	481	3.74	194.0	166.0	...	3.89	197.0	157.0	349	4.38	336.0	321.5	465	9.70	A
1	89.000000	90	95	82	306.0	249.0	576	7.81	288.0	276.5	...	8.94	282.0	278.0	464	8.54	287.5	258.0	396	8.24	C+
2	87.000000	86	85	91	250.0	161.0	502	3.96	150.0	209.0	...	6.65	172.0	184.0	305	6.46	230.0	221.5	373	8.41	B+
3	95.333333	187	199	187	206.0	149.0	430	4.78	190.0	226.5	...	7.17	175.0	197.0	309	4.79	250.5	239.0	381	7.72	B+
4	92.083333	173	194	181	285.0	238.0	552	7.07	261.5	229.5	...	7.83	248.0	225.0	405	7.90	188.0	239.0	321	7.43	C+
5	59.000000	63	54	40	264.0	218.0	543	6.96	226.5	237.0	...	7.58	204.0	251.0	388	7.48	110.5	169.0	259	2.83	F
6	95.166667	195	196	187	320.0	270.0	586	8.67	272.0	274.5	...	8.19	281.0	292.0	437	8.31	300.0	291.0	435	7.52	D
7	91.833333	185	193	185	264.0	203.0	518	6.28	200.0	221.0	...	7.19	212.0	187.0	371	5.75	241.0	191.0	385	9.20	A
8	86.166667	180	188	185	270.0	202.0	512	7.57	203.0	208.0	...	6.00	209.0	188.0	368	6.15	203.0	198.5	332	7.35	D
9	93.833333	174	185	190	300.0	254.0	550	7.93	272.5	260.0	...	7.96	235.0	241.0	392	6.92	291.5	282.5	433	8.59	A

10 rows × 21 columns

Figure 4. Sample of grade prediction dataset.

3.2. Data preprocessing and exploratory data analysis (EDA)

Data preprocessing serves as a crucial initial phase within the data analysis and machine learning pipeline. Its primary objective is to convert raw data into a well-organized, structured format that is suitable for subsequent analysis or model training. This fundamental process encompasses various tasks, including data cleansing, handling missing data, detecting outliers and performing feature engineering. Managing missing data during the data preprocessing stage holds particular importance, as it plays a pivotal role in upholding the integrity and dependability of any analytical task. The presence of missing data can introduce bias and undermine the accuracy of models, underscoring the significance of employing effective strategies to address this issue. Typically, this involves identifying missing values, making decisions regarding whether to remove or impute them, and selecting appropriate imputation methods, such as mean, median, or more advanced techniques like regression or data-driven algorithms. There are no null values present in the collected grade prediction dataset. So, there is no need for data cleaning.

Exploratory Data Analysis (EDA) is a crucial initial step in the data analysis process. Its primary goal is to visually represent, summarize, and gain a deep understanding of the essential characteristics of a dataset

	Plus 2 Percentage	Physics	Chemistry	Maths	S1_Series 1	S1_Series 2	S1_Sessional	S1_University	S2_Series 1	S2_Series 2	S2_Sessional	S2_University	S3_Serie
count	12081.000000	12081.000000	12081.000000	12081.000000	12081.000000	12081.000000	12081.000000	12081.000000	12081.000000	12081.000000	12081.000000	12081.000000	12081.000
mean	83.958043	128.471981	129.228706	123.733217	181.210661	163.367230	467.041884	6.628210	203.050733	195.685473	492.095025	7.268537	186.189
std	9.597619	46.914348	49.615556	51.934502	80.571048	74.244469	81.930256	2.072405	59.187553	57.179219	93.470483	16.637981	85.467
min	20.000000	47.000000	38.000000	33.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000
25%	78.600000	83.000000	81.000000	77.000000	111.500000	101.500000	430.000000	5.770000	167.000000	160.000000	452.000000	6.290000	108.500
50%	85.800000	137.000000	139.000000	121.000000	173.500000	155.000000	479.000000	7.270000	206.000000	198.000000	503.000000	7.460000	190.000
75%	91.500000	175.000000	180.000000	178.000000	252.000000	224.000000	519.000000	8.060000	245.000000	236.500000	549.000000	8.200000	260.500
max	115.800000	200.000000	200.000000	200.000000	406.000000	355.500000	612.000000	10.000000	329.500000	331.000000	652.000000	748.000000	416.500

Figure 5. Statistical analysis of dataset.

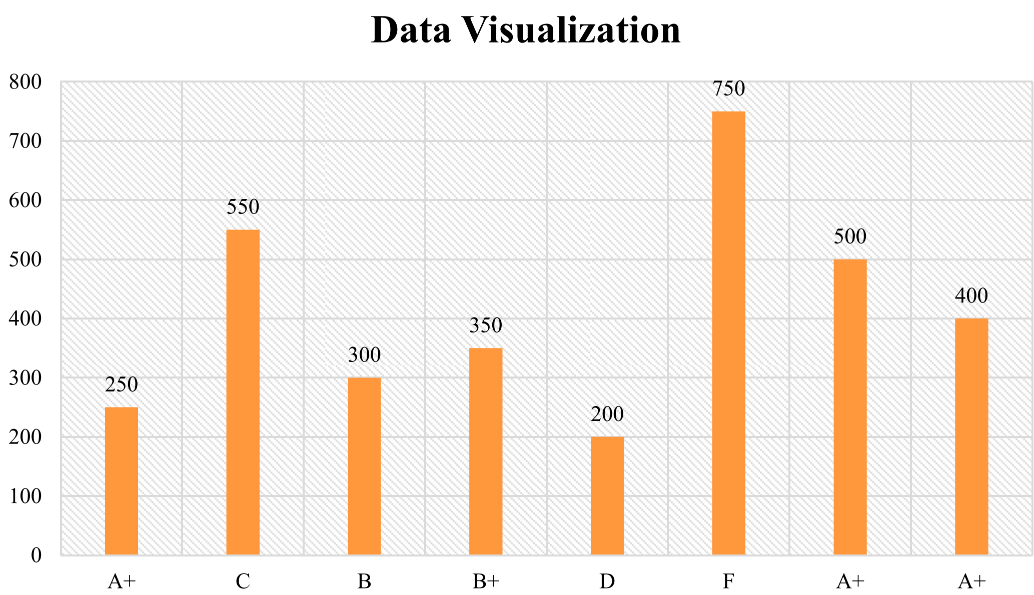

Figure 6. Data visualization of grade prediction dataset.

before delving into formal modelling or hypothesis testing. EDA techniques help in identifying data patterns, trends, anomalies, and relationships. Summary statistics play a pivotal role in simplifying and communicating the key aspects of a dataset. These summary statistics are fundamental in the early stages of data exploration, enabling further investigation and well-informed decision-making. This involves the computation of fundamental statistical measures such as the mean, median, mode, standard deviation, variance and percentiles. The statistical analysis of the collected grade prediction dataset is shown in Figure 5.

Data visualization is a crucial aspect of EDA as it converts raw data into visual representations, making complex information more accessible. This approach unveils concealed patterns and trends that might remain obscure otherwise. Typical data visualization methods include bar plots, histograms and heatmap. The graphical visualization of the counts of different grades are illustrated in Figure 6.

The histograms in Figure 7 illustrate the distribution of features, displaying the frequency of observations within each created bin. Additionally, the histograms serve the purpose of identifying potential outliers in the data.

Figure 8 represents the heatmap visualization of the grade prediction dataset. A correlation heat map is a valuable data analysis visualization tool that illustrates the associations between pairs of variables in a dataset. It is typically presented as a grid of coloured squares, where each square signifies the correlation coefficient between two variables. The colours in the squares convey both the strength and direction of the correlation: darker colours indicate stronger positive correlations, while lighter colours suggest either weaker correlations or negative relationships.

3.3. Feature optimization using analysis of variance (ANOVA)

Analysis of Variance (ANOVA) for feature optimization is a statistical method used to pinpoint the most influential features in a predictive modelling scenario, particularly when dealing with numerous input variables. It assists in the selection of a subset of features that have the greatest impact on the model's overall performance

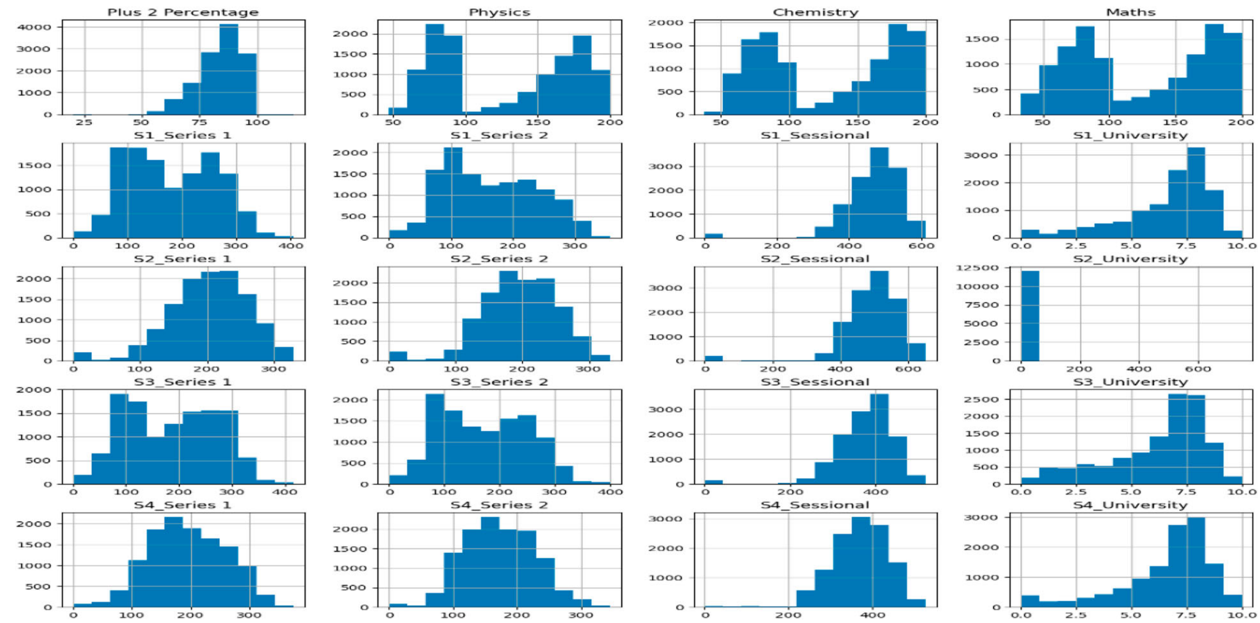

Figure 7. Feature distribution of grade prediction dataset.

[31]. ANOVA is a statistical method used to assess whether there are meaningful distinctions among the averages of different groups. It achieves this by comparing the extent of variability between these group averages to the variability within each individual group. If the disparity in variability between groups turns out to be substantially greater than the variability within the groups themselves, it implies that there are significant variations in the means of these groups. ANOVA computes an F-statistic by pitting the variability between groups against the variability within them. When the F-statistic surpasses a predetermined critical threshold, it signifies that there are indeed substantial variations in the group means. ANOVA is a hypothesis-testing approach that divides the overall variance in a dataset into two key components: the variance attributed to distinctions between the groups, and the variance stemming from differences within each group. In the context of feature optimization, ANOVA proves invaluable in discerning whether there are notable distinctions in the average values of the target variable among the various categories or levels associated with a particular feature (Figure 9).

For variance due to differences within groups:

$$\text{Sum of squares, } \mathrm{SS}_W = \sum_{j=1}^{k}\sum_{j=1}^{l}(X - \bar{X}_j)^2 \tag{1}$$

$$\text{Degrees of freedom, } \mathrm{df}_W = k - 1 \tag{2}$$

$$\text{Mean squares, } \mathrm{MS}_W = \frac{\mathrm{SS}_W}{\mathrm{df}_W} \tag{3}$$

For variance due to differences between groups:

$$\text{Sum of squares, } \mathrm{SS}_b = \sum_{j=1}^{k}(\bar{X}_j - \overline{X})^2 \tag{4}$$

$$\text{Degrees of freedom, } \mathrm{df}_b = n - k \tag{5}$$

$$\text{Mean squares, } \mathrm{MS}_b = \frac{\mathrm{SS}_b}{\mathrm{df}_b} \tag{6}$$

The F-ratio can be expressed as,

$$F = \frac{\mathrm{MS}_b}{\mathrm{MS}_w} \tag{7}$$

ANOVA involves testing two hypotheses:

(1) **Null Hypothesis (H_0):** This hypothesis assumes that there are no significant differences in the means of the target variable among the categories or levels of the feature being analysed.
(2) **Alternative Hypothesis (H_a):** This hypothesis suggests that there are statistically significant differences in the means of the target variable among the categories or levels of the feature.

ANOVA calculates an F-statistic, which represents the ratio of the variance between groups to the variance within groups. A high F-statistic indicates that there are significant differences between the groups, implying that the feature is a good candidate for feature optimization. The F-statistic is used to calculate a p-value, which quantifies the probability of observing such significant differences by random chance alone. If the p-value is below the chosen significance level, reject the null hypothesis (H_0) and conclude that the feature has a significant impact on the target variable. Based on the ranking, select a subset of the most relevant features for the predictive model. The major steps in ANOVA are explained below.

Step 1: Calculate all the means.
Step 2: Set up the null and alternate hypothesis and α.

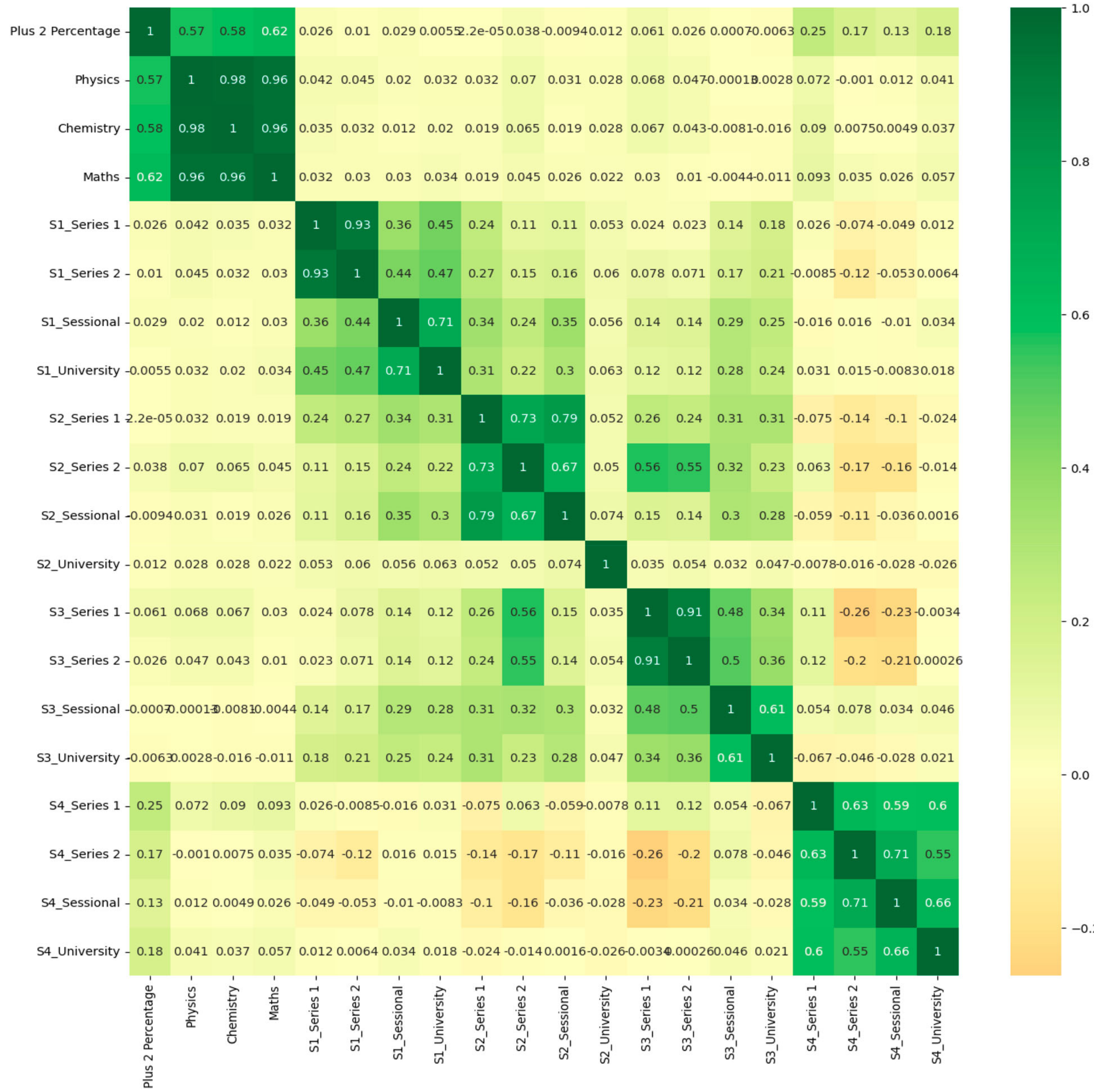

Figure 8. Heat map visualization of grade prediction dataset.

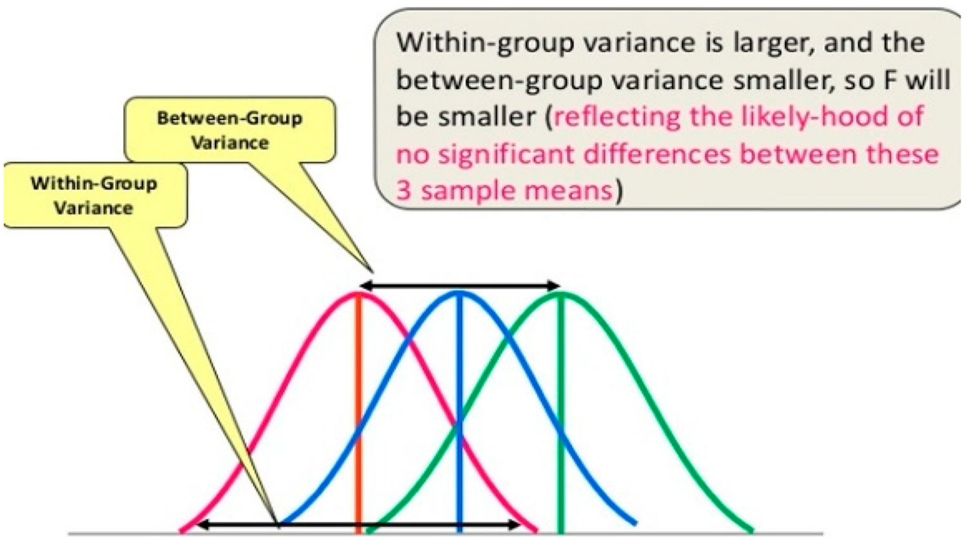

Figure 9. Source of Variation.

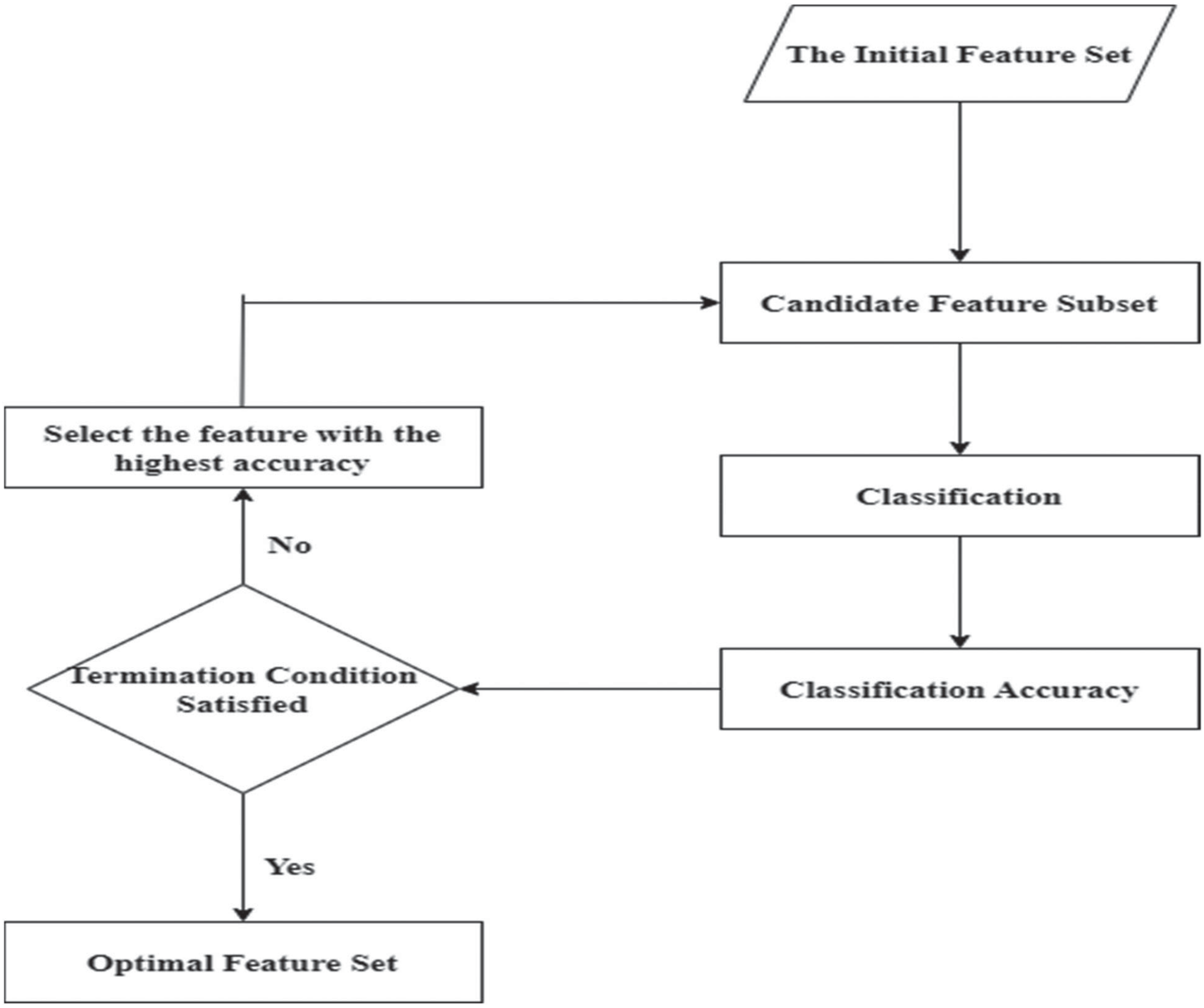

Figure 10. Flow chart of feature optimization using ANOVA.

Step 3: Calculate the Sum of Squares.
Step 4: Calculate the Degrees of Freedom (df).
Step 5: Calculate the Mean Squares.
Step 6: Create a Summary Table and Calculate the F Statistic.
Step 7: Look up F from the table.

Feature optimization using ANOVA is a valuable technique for selecting the most relevant features in a dataset by assessing their impact on a target variable. It helps in building more efficient and interpretable predictive models while ensuring statistical rigour in the feature selection process. The flow chart of feature optimization using ANOVA is depicted in Figure 10.

3.4. Grade prediction system using gated recurrent unit (GRU)

The proposed grade prediction system is based on GRU. The optimized features are fed to the GRU model for the final semester grade prediction of students. The GRU is a variation of the traditional RNN that has gained significant popularity in the field of deep learning. GRU provides a simple gating function [32]. These networks can be considered as the simplified version of LSTM model. The number of multiplicative gates is reduced to two – Update gate and Reset gate [33]. The input gate and forget gates are combined into update state and a new reset state is introduced, whose function is to moderate the impact of previous hidden state on new hidden state. The weight matrix is also reduced by one. Other simplifications in GRU are lesser tensor operations, fewer parameters. As a result, GRUs train and converge faster. Despite the simplifications, GRUs retain the property of long-term retention. GRUs perform well on Sequential as well as temporal data. The architecture of GRU is shown in Figure 11.

The GRU model is described by the following equations [34].

$$Z_t = \sigma(W_Z x_t + U_Z h_{t-1} + b_Z) \tag{8}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{9}$$

$$\hat{h}_t = \tan h(W_h x_t + U_h(h_{t-1} * r_t) + b_h) \tag{10}$$

$$h_t = Z_t * h_{t-1} + (1 - Z_t) * \hat{h}_t \tag{11}$$

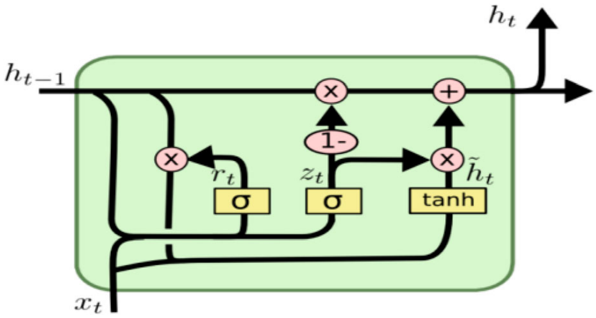

Figure 11. Basic architecture of GRU.

Table 3. Summary of proposed grade prediction system.

Layer type	Output shape	Parameters
GRU	(None, 20, 50)	7950
Dropout	(None, 20, 50)	0
GRU	(None, 20, 50)	15300
Dropout	(None, 20, 50)	0
GRU	(None, 20, 50)	15300
Dropout	(None, 20, 50)	0
GRU	(None, 50)	15300
Dropout	(None, 50)	0
Dense	(None, 10)	510
Total Parameters: 54,360		
Trainable Parameters: 54,360		
Non- Trainable Parameters: 0		

where Z_t = Vector corresponding to the update gate. It determines how much of the past knowledge needs to be passed along into the future. It is analogous to the Output Gate in an LSTM network.

r_t = Vector corresponding to the reset gate. It determines how much of the past knowledge to forget. It is analogous to the combination of the Input Gate and the Forget Gate in an LSTM recurrent unit.

h_t = State vector for current time frame t. It is often overlooked during a typical discussion on GRU Network. It is incorporated into the Reset Gate just like the Input Modulation Gate is a sub-part of the Input Gate and is used to introduce some non-linearity into the input and to also make the input Zero-mean. Another reason to make it a sub-part of the Reset gate is to reduce the effect that previous information has on the current information that is being passed into the future.

The architecture of the proposed grade prediction model is shown in Figure 12. The proposed grade prediction model consists of four GRU layers. A dropout layer is placed after each GRU layer. Dropout is a regularization technique that helps prevent overfitting by randomly setting a fraction of input units to zero during training. The final layer is a dense layer with 10 units and a SoftMax activation function. This is the output layer that produces probability distributions over 10 classes (grades).

The model summary of the proposed grade prediction model is tabulated in Table 3.

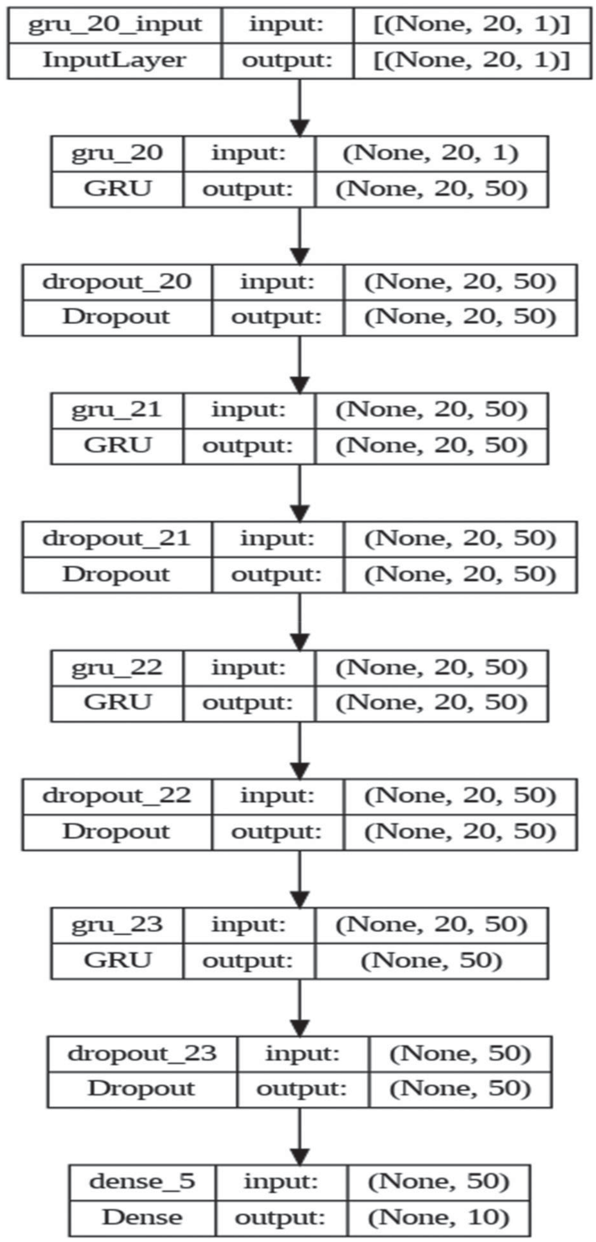

Figure 12. Model architecture of proposed grade prediction system.

4. Results and discussion

4.1. *Hardware and software setup*

The proposed system utilizes handcrafted dataset contains 21 attributes. Google Colaboratory and Microsoft windows 10 are chosen in this research to ensure a stable computing experience. The system is equipped with an Intel Core i7-6850 K 3.60 GHz 12 – core processor, one NVIDIA GeForce GTX 1080 Ti GPU 2760 4 MB. The dataset is divided into two sets: training set (80%) and test set (20%). The entire procedure made use of Python and TensorFlow. The "Adam" optimization function was used in the proposed model. The categorical crossentropy is used as the loss function. The model was trained using batch size of 128 across 500 epochs. Finally, the proposed model predicts the

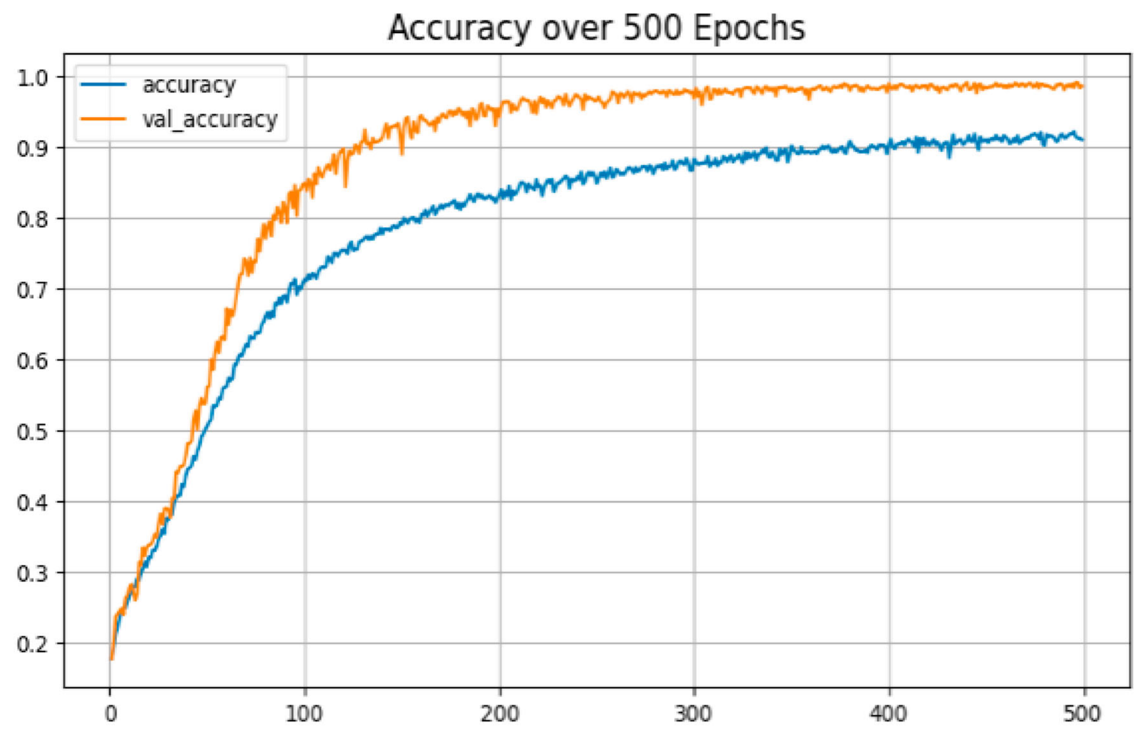

Figure 13. Accuracy plot of proposed grade prediction system.

grade of the students to either of the 10 grades S, A+, A, B, B+, C, C+, D, P, F.

4.2. Performance evaluation

The proposed grade prediction system is evaluated on generally available performance metrics such as accuracy, precision, recall and F1-score. Accuracy quantifies the percentage of correctly categorized instances among all the instances within the dataset.

$$\text{Accuracy} = \frac{(\text{TP} + \text{FP})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \tag{12}$$

Precision quantifies how many of the positive predictions made by the model were actually correct.

$$\text{Precision} = \frac{(\text{TP})}{(\text{TP} + \text{FP})} \tag{13}$$

Recall, also known as Sensitivity or True Positive Rate, measures the proportion of actual positive instances that were correctly predicted as positive by the model. It focuses on the ability of the model to capture all positive instances.

$$\text{Recall} = \frac{(\text{TP})}{(\text{TP} + \text{FN})} \tag{14}$$

The F1-score is the harmonic mean of precision and recall. It balances the trade-off between precision and recall, providing a single metric that considers both false positives and false negatives.

Table 4. Performance evaluation of proposed grade prediction system.

Performance metrics	Obtained results
Accuracy Score	98.95%
Misclassification Score	1.05
Precision Score	98.94%
Recall Score	98.93%
F1-Score	98.94%

$$F1 - \text{Score} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{15}$$

The classification report of the proposed grade prediction model after simulation is tabulated in Table 4.

Accuracy and loss plots are crucial resources for assessing and tracking the progress of training and evaluating machine learning and deep learning models. These visualizations offer valuable insights into a model's capacity to learn from the training data and its performance on validation or test data. The accuracy plot illustrates the evolution of a model's prediction accuracy across epochs during training, quantifying the fraction of correctly predicted instances relative to the total instances in the validation or test dataset. The accuracy plot of the proposed model is illustrated in Figure 13.

The loss plot is a graphical representation that illustrates how the model's loss evolves across training epochs. Loss, in this context, is a single numeric value that measures how far off the model's predictions are from the true target values. When the loss curve consistently decreases over epochs, it signifies that the model

Figure 14. Loss plot of proposed grade prediction system.

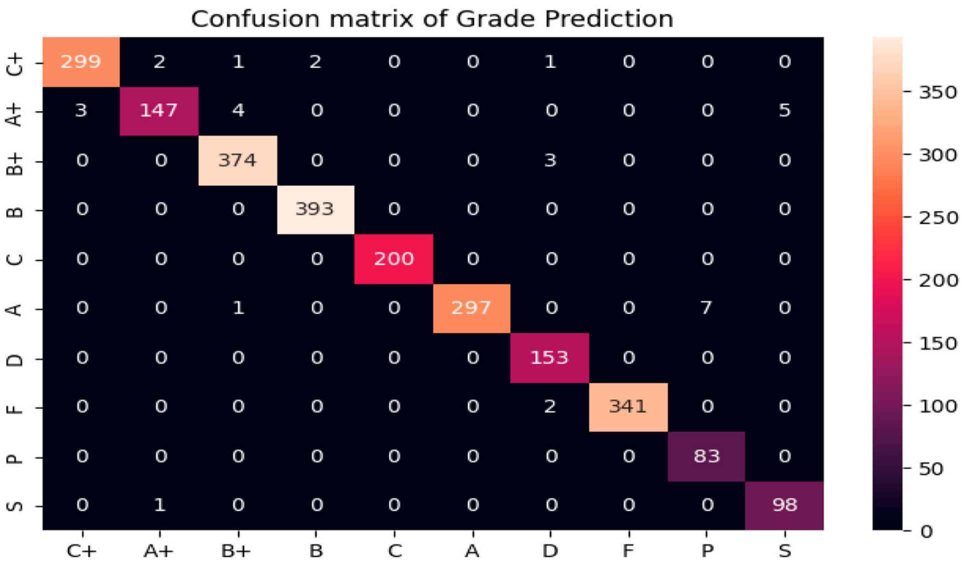

Figure 15. Confusion matrix for grade prediction system.

is progressively improving its prediction accuracy as it becomes more familiar with the training data. The loss plot of proposed grade prediction system is visualized in Figure 14.

A confusion matrix is a fundamental tool for evaluating the performance of a model, providing a detailed breakdown of the model's predictions and their correspondence with the actual ground truth. It's particularly useful when assessing how well a model performs across different classes. The confusion matrix obtained for proposed grade prediction system is shown in Figure 15.

The proposed model for grade prediction shows amazing prediction in terms of performance parameters. The proposed method can effectively be used for final semester grade prediction of students. Some of the prediction outputs obtained using the proposed prediction system is tabulated in Table 5.

4.3. Comparison

The performance of the proposed grade prediction system is compared with another grade prediction systems which utilized Convolutional Neural Network with

Table 5. Random prediction outputs.

Input data	Predicted grade
[96.6,93,97,95,222,189,481,3.74,194,166,460,3.69,197,157,349,4.38,336,321.5,465,9.7]	A
[69,69,79,55,260,209,520,7.78,281.5,232,505,7.77,231,268,409,7.38,160,137.5,271,3.7]	F
[88.08333333,160,180,165,317,254,591,8.78,295,279,531,8.5,282,264,448,8.58,223.5,187,321,5.13,]	C+
[94.16666667,186,194,193,150,145,420,2.61,148.5,141,373,3.29,160,97,288,5.04,215,168.5,312,6.09]	D
[94.5,179,181,200,215,194,497,6.61,147.5,191,445,5.83,156,183,330,4.42,164.5,141,271,4.61]	P

Table 6. Performance comparison of grade prediction systems.

	GRU with feature Optimization using analysis of variance (ANOVA)	Long short-term memory (LSTM) with feature Optimization using L1-Norm	Convolutional neural network with feature Optimization using Bayesian optimizer
Accuracy score	98.95%	97.87%	98.38%
Precision score	98.94%	97.88%	98.39%
Recall score	98.93%	97.87%	98.38%
F1-score	98.94%	97.87%	98.38%
Misclassification score	1.05	1.87	1.6135

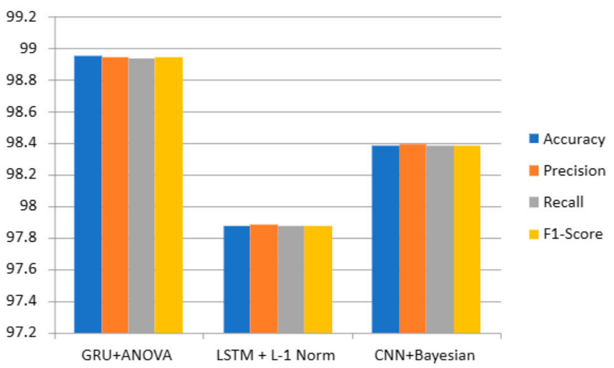

Figure 16. Performance comparison of proposed grade prediction models.

Bayesian Optimization and Long Short-Term Memory with L1-Norm Optimization. The performance comparison is tabulated in Table 6.

The proposed grade prediction system using GRU with ANOVA outperformed the other two models with an accuracy score of 98.95%, precision score of 98.94%, recall score of 98.93% and f1-score of 98.94% with a misclassification score of 1.05. The simulation results revealed that the proposed grade prediction model can be effectively be used for final semester grade prediction of students. The graphical visualization of performance comparison is shown in Figure 16.

The comprehensive summaries of the results attained from the experiments are discussed below:

- Final semester grade prediction system using GRU with feature optimization using ANOVA is mainly discussed in this paper.
- The handcrafted grade prediction dataset was utilized in this work, which consists of 12,081 rows and 21 columns.
- Feature optimization is done with the help of Analysis of Variance (ANOVA). The final grades are predicted with the help of GRU-based multiclass classifier.
- The simulation results revealed that the GRU-based grade prediction system with ANOVA outperformed the grade prediction system using LSTM+L1-Norm and CNN + Bayesian Optimization with 98.95% accuracy score.

5. Conclusion

The prediction of final semester grades for students is a crucial tool in assessing their academic performance and identifying the likelihood of failing. By analyzing variables, educators and institutions can gain valuable insights into a student's progress and potential outcomes. Predictive models and data analysis techniques have proven to be effective in providing early warnings about students who may be at risk of failing. These tools not only help educators intervene and provide additional support to struggling students but also allow institutions to allocate resources more efficiently to address the specific needs of their student body. This paper proposes an effective deep learning based model for final semester grade prediction using GRU and the features are optimized using ANOVA. The predictive model is trained and evaluated on a handcrafted grade prediction dataset, which contains the academic performance of the students during Plus 2 and from Semester 1 to Semester 4. The use of GRU models in conjunction with feature optimization through ANOVA holds great promise for predicting the final semester grades of students and assessing their academic performance. The application of GRU models in this context allows us to leverage the power of recurrent neural networks, which excel at capturing sequential patterns and dependencies in data. Furthermore, the integration of feature optimization using ANOVA helps refine the predictive model by selecting the most relevant features from the dataset. The simulation results demonstrated that the proposed model achieved better prediction performance with an accuracy score of 99.54% and outperformed the grade prediction systems using CNN with Bayesian Optimizer and LSTM with L1-Norm Optimizer. This data-driven methodology has the potential to transform the way educational institutions support their students, fostering a culture of academic success and achievement.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] Gupta MK, Chandra P. A comprehensive survey of data mining. Int J Inf Technol . 2020;12(4):1243–1257.
[2] Larose DT. Discovering knowledge in data: An introduction to data mining. New Delhi: John Wiley & Sons, Inc; 2014.
[3] Posselt JR, Grodsky E. Graduate education and social stratification. Annu Rev Sociol. 2017;43:353–378. doi: 10.1146/annurev-soc-081715-074324
[4] Agarwal P. (2006). Higher education in India: The need for change (No. 180). Working paper.
[5] Aud S, Wilkinson-Flicker S. The Condition of Education 2013. U.S. Department of Education, National Center for Education Statistics., 2013, no. NCES 2013-037.
[6] N. R. Council, Building a workforce for the information economy. US: National Academies Press, 2001.
[7] Sreenivasulu MD, Devi JS, Arulprakash P, et al. Implementation of latest machine learning approaches for students grade prediction. Int J Early Child. 2022;14(3):3027–3057.
[8] Bujang SDA, Selamat A, Ibrahim R, et al. Multiclass prediction model for student grade prediction using machine learning. IEEE Access. 2021;9:95608–95621. doi:10.1109/ACCESS.2021.3093563
[9] Zhang Y, Yun Y, Dai H, et al. Graphs regularized robust matrix factorization and its application on student grade prediction. Applied Sciences. 2020;10(5):1755. doi:10.3390/app10051755
[10] Kanetaki Z, Stergiou C, Bekas G, et al. Grade prediction modeling in hybrid learning environments for sustainable engineering education. Sustainability. 2022;14(9):5205. doi:10.3390/su14095205
[11] Zhang Y, An R, Cui J, et al. Undergraduate grade prediction in Chinese higher education using convolutional neural networks. In: Lak21: 11th International Learning Analytics and Knowledge Conference. Germany. 2021 Apr. p. 462–468.
[12] Damuluri S, Islam K, Ahmadi P, et al. Analyzing navigational data and predicting student grades using support vector machine. Emerg Sci J. 2020;4(4):243–252. doi:10.28991/esj-2020-01227
[13] Ha DT, Loan PTT, Giap CN, et al. An empirical study for student academic performance prediction using machine learning techniques. Int J Comp Sci Inf Secur. 2020;18(3):75–82.
[14] Masangu L, Jadhav A, Ajoodha R. Predicting student academic performance using data mining techniques. advances in science. Technol Eng Syst J. 2021;6(1):153–163.
[15] Obsie EY, Adem SA. Prediction of student academic performance using neural network, linear regression and support vector regression: a case study. Int J Comput Appl. 2018;180(40):39–47.
[16] Alhassan A, Zafar B, Mueen A. Predict students' academic performance based on their assessment grades and online activity data. Int J Adv Comp Sci Appl. 2020; 11(4):185–192. doi:10.14569/IJACSA.2020.0110425
[17] Naicker N, Adeliyi T, Wing J. Linear support vector machines for prediction of student performance in school-based education. Math Probl Eng. 2020;2020: 1–7. doi:10.1155/2020/4761468
[18] Alshabandar R, Hussain A, Keight R, et al. Students performance prediction in online courses using machine learning algorithms. In: 2020 International Joint Conference on Neural Networks (IJCNN). United Kingdom: IEEE; 2020 Jul. p. 1–7.
[19] Alamri LH, Almuslim RS, Alotibi MS, et al. Predicting student academic performance using support vector machine and random forest. In: Proceedings of the 2020 3rd International Conference on Education Technology Management. New York: Association for Computing Machinery; 2020 Dec. p. 100–107.
[20] Bhutto ES, Siddiqui IF, Arain QA, et al. Predicting students' academic performance through supervised machine learning. In: 2020 International Conference on Information Science and Communication Technology (ICISCT. Pakistan: IEEE; 2020 Feb. p. 1–6.
[21] Rai S, Shastry KA, Pratap S, et al. Machine learning approach for student academic performance prediction. In: Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020). Vol. 1. Singapore: Springer; 2021. p. 611–618.
[22] Lim TW, Khor KC, Ng KH. Dimensionality reduction for predicting student performance in unbalanced data sets. Int J Advance Soft Compu Appl. 2019;11(2): 76–87.
[23] Nabil A, Seyam M, Abou-Elfetouh A. Prediction of students' academic performance based on courses' grades using deep neural networks. IEEE Access. 2021; 9:140731–140746. doi:10.1109/ACCESS.2021.3119596
[24] Saifudin A, Desyani T. Forward selection technique to choose the best features in prediction of student academic performance based on Naïve Bayes. J Phys Conf Ser. 2020 March;1477(3):032007. IOP Publishing. doi:10.1088/1742-6596/1477/3/032007
[25] Mubarak AA, Cao H, Hezam IM. Deep analytic model for student dropout prediction in massive open online courses. Comput Electr Eng. 2021;93:107271. New York. doi:10.1016/j.compeleceng.2021.107271
[26] Bouchard K, Gonzales L, Maitre J, et al. Features exploration for grades prediction using machine learning. In: Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good. Portugal. 2020 Sep. p. 78–83.
[27] Ayienda R, Rimiru R, Cheruiyot W. Predicting students' academic performance using a hybrid of machine learning algorithms. In: 2021 IEEE AFRICON. Kenya: IEEE; 2021 Sep. p. 1–6.
[28] Zhang X, Xue R, Liu B, et al. Grade prediction of student academic performance with multiple classification models. In: Proc.14th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNCFSKD). China. Jul. 2018, p. 1086–1090.
[29] Deepika K, Sathyanarayana N. Hybrid model for improving student academic performance. Int J Adv Res Eng Technol. 2020;11(10):768–779.
[30] Ashraf M, Zaman M, Ahmed M. An intelligent prediction system for educational data mining based on ensemble and filtering approaches. Procedia Comput Sci. 2020;167:1471–1483. doi:10.1016/j.procs.2020.03.358

[31] Kim TK. Understanding one-way ANOVA using conceptual figures. Korean J Anesthesiol. 2017;70(1):22–26. doi:10.4097/kjae.2017.70.1.22

[32] Chung J, Gulcehre C, Cho K, et al. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* preprint arXiv:1412.3555.

[33] Dey R., Salem FM. Gate-variants of gated recurrent unit (GRU) neural networks. In: 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS). North America: IEEE; 2017 Aug, p. 1597–1600.

[34] Ravanelli M, Brakel P, Omologo M, et al. Light gated recurrent units for speech recognition. IEEE Trans Emerg Topics Comput Intellig. 2018;2(2):92–102. doi:10.1109/TETCI.2017.2762739