BULLETIN OF EDUCATION AND PSYCHOLOGY
Research Article | Volume 5 | Issue 5 | Mar 2025
Received: 22 Feb 2025 | Revised: 29 Feb 2025
Accepted: 26 Feb 2025 | Published Online: 19 Mar 2025

Prediction of Student Disciplinary Behavior through Efficient Ridge Regression

Zhiqiang Zhao [1,*], Ping Ren[2]

[1] Beijing PhD Village Education Technology Co., Ltd; Beijing 100871, China

[2] Chengdu Ding Yi Education Consulting Co., Ltd, Chengdu 610023, China

[*]Corresponding Author, Email: 202128030258@mail.bnu.edu.cn

Abstract: In light of the importance of predicting student disciplinary behavior in educational settings, this paper addresses the current challenges and limitations in existing research methodologies. The ability to accurately anticipate disciplinary issues is crucial for maintaining a safe and conducive learning environment. However, traditional prediction models often lack efficiency and effectiveness in this context. To overcome these limitations, this paper proposes a novel approach utilizing efficient Ridge Regression for the prediction of student disciplinary behavior. By integrating this innovative technique with relevant behavioral data, our study aims to provide a more precise and reliable predictive model for identifying at-risk students and implementing targeted intervention strategies. This research contributes to the advancement of predictive analytics in education and underscores the significance of proactive measures in managing disciplinary challenges.

Keywords: *Disciplinary Behavior; Prediction Models; Ridge Regression; Predictive Analytics; Intervention Strategies*

1. Introduction

The field of Student Disciplinary Behavior focuses on studying and analyzing the behavior of students in educational settings, with the goal of understanding, predicting, and addressing disciplinary issues. Current challenges and bottlenecks in this field include the limited availability of comprehensive data on student behavior, the complexity of factors influencing disciplinary incidents, and the need for more effective strategies for prevention and intervention. Additionally, there is a lack of standardized definitions and measurement tools for disciplinary behavior, making it difficult to compare research findings across studies. These challenges highlight the importance of continued research and collaboration in order to improve our understanding of student disciplinary behavior and develop evidence-based practices for promoting positive behavior and preventing misconduct.

To this end, current research on student disciplinary behavior has advanced to encompass a multidisciplinary approach, integrating insights from psychology, sociology, education, and criminology. Studies now focus on not only understanding the causes and consequences of disciplinary issues but also on developing effective intervention strategies for promoting positive behavior in educational settings. A literature review was conducted to explore various aspects of educators' roles in managing students' disciplinary behavior through counseling guidance programs. Jannah [1] and Jannah [2] investigated educators' understanding of such programs, strategies applied, and perceptions of effectiveness. McCormack et al. [3] linked temperature extremes with absenteeism and disciplinary issues, highlighting environmental influences. Muslida et al. [4] focused on exemplary teacher behavior impacting student disciplinary behavior. Erwin et al. [5] studied the effects of increased recess time on disciplinary referrals and academic performance. Gennetian et al. [6] examined SNAP benefit cycles' association with disciplinary infractions. Manalo [7] addressed disciplinary issues in a Catholic private school. Kunesh and Noltemeyer [8] explored pre-service teachers' perceptions of black boys' behavior stereotypes. Magier et al. [9] analyzed school policies and disciplinary consequences on student cannabis use. Gage et al. [10] replicated SWPBIS's impact on disciplinary exclusions, emphasizing the positive effects on reducing suspensions. These studies collectively contribute to understanding educators' crucial role in addressing students' disciplinary behavior within different contexts. A literature review explored educators' roles in managing students' disciplinary behavior through counseling guidance programs. Ridge Regression is essential for analyzing the complex relationships between various factors impacting disciplinary issues. Its regularization technique helps prevent overfitting and provides more stable estimates, making it a suitable choice for modeling the multidimensional nature of disciplinary behavior research.

Specifically, Ridge Regression can be utilized to analyze the impact of various factors on student disciplinary behavior. By incorporating regularization techniques, Ridge Regression can effectively handle multicollinearity issues in the data, providing more accurate and reliable results when studying the predictors of disciplinary incidents in educational settings. Ridge regression has been studied as a method of biased estimation for nonorthogonal problems [11]. The saturation effect of kernel ridge regression has been a long-standing conjecture, and recent research has provided a proof of this phenomenon [12]. Furthermore, the application of kernel ridge regression in graph dataset distillation has led to the development of efficient models such as Kernel Ridge Regression-Based Graph Dataset Distillation (KIDD) [13]. Another study has addressed the optimality of misspecified kernel ridge regression for different smoothness levels of the true function [14]. Additionally, dataset meta-learning from kernel ridge-regression has introduced the concept of ϵ-approximation of datasets using Kernel Inducing Points (KIP) and has shown significant improvements in dataset distillation [15]. Moreover, the use of ridge regression and other kernels for genomic selection has been facilitated by the development of the R package rrBLUP [16]. Random matrix theory has been applied to ridge regression in high-dimensional statistics, providing non-asymptotic bounds for bias and variance in a dimension-free framework [17]. Finally, the ensemble of machine learning models with ridge regression has been applied to solar and wind forecasting, demonstrating its effectiveness in practical applications [18]. However, current limitations include the need for further exploration of the generalizability of kernel ridge regression in diverse datasets, potential challenges in scaling up the use of Kernel Inducing Points (KIP) for dataset meta-learning, and the requirement for more comprehensive benchmarks to evaluate the performance of ridge regression in genomic selection.

To overcome those limitations, this paper aims to address the current challenges in predicting student disciplinary behavior in educational settings. The importance of accurately foreseeing disciplinary issues for creating a safe learning environment is emphasized, yet traditional prediction models have shown inefficiency and ineffectiveness. In response, this study proposes a novel approach that leverages Ridge Regression as an efficient method for predicting student disciplinary behavior. This innovative technique is integrated with behavioral data to enhance the precision and reliability of the predictive model. By focusing on identifying at-risk students and developing targeted intervention strategies, the research endeavors to contribute to the advancement of predictive analytics in education. The emphasis on proactive measures underscores the significance of early intervention in managing disciplinary challenges, highlighting the potential impact of this research on improving educational outcomes.

This study delves into the pressing issue of predicting student disciplinary behavior in educational settings. Section 2 elaborates on the problem statement, highlighting the challenges and limitations in existing research methodologies. Section 3 introduces the proposed method, emphasizing the utilization of efficient Ridge Regression for prediction. In Section 4, a case study is presented to illustrate the application of the novel approach. The analysis of results in Section 5 demonstrates the effectiveness of the predictive model in identifying at-risk students. Section 6 delves into a comprehensive discussion, emphasizing the significance of proactive measures in managing disciplinary challenges. Finally, Section 7 offers a succinct summary, showcasing the research's contribution to advancing predictive analytics in education and its potential for implementing targeted intervention strategies.

2. Background

2.1 Student Disciplinary Behavior

Student Disciplinary Behavior (SDB) refers to the actions, practices, and measures taken by educational institutions to manage student conduct and ensure adherence to established codes of behavior. The evaluation of such behavior utilizes diverse methodologies combining qualitative and quantitative assessments, providing insight into the impact of discipline on educational outcomes.

To effectively model SDB, we explore parameters such as violation frequency, λ, disciplinary actions, δ, and rehabilitation rate, ρ. Each parameter elucidates different dimensions of student behavior within an educational context. Firstly, let's define the frequency of violations, represented as λ. It quantifies the average occurrences of disciplinary infractions per unit of time.

$$\lambda = \frac{N_v}{T} \tag{1}$$

where N_v is the total number of violations and T represents the time period under consideration. This formula serves as a foundational metric for understanding the prevalence of student infractions. The severity of disciplinary measures, denoted as δ, describes the intensity of the punitive actions administered upon a student's violation. The measurement includes various factors, such as detention hours, suspension periods, or expulsion:

$$\delta = \sum_{i=1}^{k} w_i x_i \tag{2}$$

Here, x_i denotes individual disciplinary measures, while w_i is a weight reflecting the severity attributed to each measure. Rehabilitation rate, ρ, embodies the capability of corrective measures to amend students' behavior toward expected norms. It can be described by the efficacy of post-discipline interventions:

$$\rho = \frac{S_r}{S_d} \tag{3}$$

where S_r represents successfully rehabilitated students, and S_d is the total number of disciplined students. A higher ρ indicates effective disciplinary processes promoting positive behavioral change. Next, the overall effectiveness of disciplinary actions, E, can be modeled as a combination of behavioral correction and prevention measures:

$$E = \alpha\lambda + \beta\rho \tag{4}$$

with α and β being coefficients reflecting the societal or institutional emphasis on the frequency and rehabilitation effectiveness, respectively, modulating these effects upon overall student behavior management. While individual models offer insights into distinct components of SDB, integrating them provides a holistic perspective. We can introduce a composite index for evaluating disciplinary performance, I_d, as follows:

$$I_d = \gamma \cdot \delta + \epsilon \cdot (1 - \rho) \tag{5}$$

where γ and ϵ represent significance attached to the severity and effectiveness of rehabilitative processes, respectively. Furthermore, considering external influences like peer pressure, socio-economic factors, and institutional integrity, an extended model extends the SDB framework:

$$SDB = f(\lambda, \delta, \rho, \sigma) \tag{6}$$

where σ aggregates external factors influencing behavior. Incorporating these allows for a comprehensive understanding of dynamics at play beyond the immediate educational setting. In conclusion, Student Disciplinary Behavior encapsulates a complex array of actions aiming at maintaining order and promoting ethical conduct. Through rigorous modeling, integrating both quantitative measures and contextual nuances, stakeholders can better gauge the effectiveness of disciplinary policies, paving the path for refined strategies and improved educational climates.

2.2 Methodologies & Limitations

The field of Student Disciplinary Behavior (SDB) has seen the deployment of various advanced methodologies for assessing and predicting disciplinary outcomes. These methodologies often incorporate both statistical and algorithmic approaches, aiming to provide a comprehensive understanding of how disciplinary measures impact student behavior. One common methodological approach is the utilization of regression analysis to predict outcomes based on disciplinary actions. This model typically involves a predictive variable for the probability of a student reoffending, P_R . Logistic regression can be employed where:

$$P_R = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \lambda + \beta_2 \delta + \beta_3 \rho)}} \tag{7}$$

Here, β_0 is the intercept, while β_1, β_2, and β_3 are the coefficients associated with violation frequency λ , severity of disciplinary measures δ , and rehabilitation rate ρ respectively. This model helps in assessing the likelihood of future infractions based on past behavior and disciplinary actions. Another contemporary method is Markov chains, which model transitions between various states of student behavior. The transition matrix, P , represents probabilities of moving from one state to another, such as from compliant, C , to non-compliant, NC , as influenced by disciplinary measures:

$$P = \begin{bmatrix} P(C \to C) & P(C \to NC) \\ P(NC \to C) & P(NC \to NC) \end{bmatrix} \tag{8}$$

This offers an insightful view into how disciplinary factors encourage persistence in compliance or recidivism. Machine learning algorithms have also risen in prominence, particularly for identifying patterns and predicting behavior. For instance, decision trees are often used to classify the outcomes of disciplinary interventions:

$$G_i(T) = 1 - \sum_j p_j^2 \tag{9}$$

where $G_i(T)$ represents the Gini impurity of node i in a decision tree, and p_j is the proportion of samples in node i belonging to class j . This can help schools decide optimal disciplinary actions based on predicted outcomes. Despite the sophistication of these methods, there are notable deficiencies. One significant issue is the potential bias in data collection and processing. Data on student behavior and disciplinary actions may not fully capture informal or unreported incidents, as well as variances in enforcement among different educational institutions. Another concern is the over-reliance on quantifiable measures. While metrics like λ , δ , and ρ provide valuable data, they do not account for the nuanced social and psychological factors affecting student behavior, such as mental health, personal circumstances, or cultural influences.

Moreover, ethical considerations arise with predictive models, particularly regarding how predictions might influence disciplinary decisions about individuals. The risk of reinforcing existing disparities or prejudices if predictions are inappropriately weighted or interpreted is a serious concern. In response to these limitations, recent research emphasizes the need to combine quantitative methods with qualitative insights, integrating student interviews, observational studies, and contextual analyses into SDB evaluations. By doing so, educational institutions can adopt more holistic and equitable strategies, ensuring that disciplinary processes not only prevent infractions but also support the developmental needs of students. To encapsulate the entire disciplinary framework in a sophisticated manner, we propose the following compound equation integrating both quantitative measures and broader contextual variables:

$$SDB_{eff} = \omega(\alpha\lambda + \beta\rho) + \nu \cdot \Theta(C, S, E) \tag{10}$$

where SDB_{eff} is the effective measure of SDB incorporating quantitative impacts and qualitative context, ω is the weighting factor for measurable impacts, and $\nu \cdot \Theta(C, S, E)$ represents the qualitative context viewed through factors like culture C, socio-economic factors S, and educational environment E. This model endeavors to bridge the gap between numerical evaluation and real-world complexity.

3. The proposed method

3.1 Ridge Regression

Ridge Regression is a widely used technique in the landscape of regression analysis, especially when dealing with multicollinearity or when the dataset's number of predictors exceeds the number of observations. Standard linear regression seeks to minimize the sum of squared residuals:

$$RSS = \sum_{i=1}^{n} (y_i - y_i)^2 \tag{11}$$

where y_i denotes the observed values and $\hat{y}_i$ denotes the predicted values. When multicollinearity is present, the variance of the coefficient estimates can become excessively large, leading to overfitting of the model. To mitigate this issue, Ridge Regression introduces a regularization term to the loss function. The modified objective is to minimize:

$$RSS + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{12}$$

Here, the additional term $\lambda \sum_{j=1}^{p} \beta_j^2$ is the ridge penalty, where λ is a non-negative tuning parameter and p is the number of predictors. The coefficients β_j represent the strength and direction of the relationship between the independent variables and the dependent variable. The aim is to shrink the coefficients towards zero but not to exactly zero, depending on λ 's magnitude. In Ridge Regression, the estimated coefficients $\hat{\beta}_{\text{ridge}}$ are obtained by solving the following equation, which incorporates the penalty term:

$$\beta_{\text{ridge}} = (X^T X + \lambda I)^{-1} X^T y \tag{13}$$

where X is the design matrix of the predictors, I is the identity matrix, and y is the response vector. The regularization term λI is crucial as it adds a diagonal matrix to $X^T X$, making it invertible even if $X^T X$ is singular, thus stabilizing the solution. The choice of λ is critical and is usually determined via cross-validation. As λ increases, the flexibility of the model decreases, leading to biased estimates with reduced variance. The balance between bias and variance is central to regularization techniques like Ridge Regression. The effect is visualized by examining the impact of λ on the estimated coefficients:

$$\lim_{\lambda\to 0}\beta_{\text{ridge}} = \beta_{\text{OLS}} \tag{14}$$

$$\lim_{\lambda\to\infty}\beta_{\text{ridge}} = 0 \tag{15}$$

This demonstrates that with $\lambda = 0$, Ridge Regression reduces to ordinary least squares (OLS), while as λ approaches infinity, it heavily regularizes the coefficients, reducing them towards zero but generally never reaching exactly zero unless the penalty is extreme. A primal understanding of Ridge Regression can be encapsulated in the objective function's transformation into a matrix form, often written for computational convenience:

$$L = \|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2 \tag{16}$$

where $\|\cdot\|_2$ represents the Euclidean norm. This expression succinctly captures the dual terms that Ridge Regression simultaneously optimizes. Consider a practical scenario where Ridge Regression is invaluable: a situation with near-collinear predictors causing variance inflation in OLS estimates. Ridge Regression circumvents this by imposing a penalty on the coefficient's size, thus preference is naturally given to solutions with smaller norms. This characteristic makes Ridge Regression particularly robust in high-dimensional settings or those plagued by multicollinearity. In summary, Ridge Regression is a key statistical technique that extends the capability of linear models by introducing regularization. This approach not only assuages the impacts of multicollinearity but also provides a flexible framework to improve generalization by adjusting the complexity of the model through the parameter λ . Understanding and applying this technique empowers researchers and data scientists to enhance the robustness and predictive accuracy of regression models in complex data landscapes.

3.2 The Proposed Framework

In understanding Student Disciplinary Behavior (SDB), it becomes imperative to utilize systematic methodologies that allow for quantifiable assessment and predictive analytics. One promising method is Ridge Regression, which effectively addresses issues of multicollinearity among various predictors integral to SDB modeling. The interplay between SDB parameters, specifically violation frequency λ , severity of disciplinary actions δ , and rehabilitation rate ρ , integrates profoundly with Ridge Regression. We first establish foundational concepts in SDB. The violation frequency is defined as follows:

$$\lambda = \frac{N_v}{T} \tag{17}$$

where N_v indicates the total number of violations while T represents the observational time frame. This quantifies the incidences of disciplinary behavior over time. To complement this, we define the severity of disciplinary actions:

$$\delta = \sum_{i=1}^{k} w_i x_i \tag{18}$$

In this equation, x_i refers to individual measures of discipline, and w_i denotes the corresponding weights reflecting their severity. The rehabilitation rate, critical for assessing intervention success, is expressed as:

$$\rho = \frac{S_r}{S_d} \tag{19}$$

Here, S_r is the number of successfully rehabilitated students, while S_d represents the total number of disciplined students, thus providing a measure of the efficacy of corrective action. When modeling the overall effectiveness of these parameters in predicting SDB, we can formulate:

$$E = \alpha\lambda + \beta\rho \tag{20}$$

The coefficients α and β highlight institutional priorities regarding the frequency of violations and rehabilitation efficiency respectively. To gain a composite index for disciplinary performance denoted as I_d , we encapsulate both severity and rehabilitative effectiveness:

$$I_d = \gamma \cdot \delta + \epsilon \cdot (1 - \rho) \tag{21}$$

where γ and ϵ are weights signifying the importance of severity and rehabilitation effectiveness. In applying Ridge Regression to the SDB framework, we aim to model an outcome variable, potentially linked to student behavior, as a function of these predictors. The standard regression aims to minimize the residual sum of squares given by:

$$RSS = \sum_{i=1}^{n} (y_i - y_i)^2 \tag{22}$$

In the presence of multicollinearity among the predictors (which could include λ , δ , and ρ among others), Ridge Regression modifies this objective function to include a regularization term:

$$RSS + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{23}$$

This additional term, characterized by $\lambda \sum_{j=1}^{p} \beta_j^2$, acts as a penalty that shrinks the coefficients towards zero, minimizing the impact of multicollinearity. The resulting coefficient estimates obtained through Ridge Regression are calculated via:

$$\beta_{\text{ridge}} = (X^T X + \lambda I)^{-1} X^T y \tag{24}$$

In this equation, X is the design matrix representing our predictors (e.g., λ , δ , and ρ), y is the response vector of interest, and I is the identity matrix, which plays a crucial role in ensuring the matrix inversion remains stable, thereby mitigating issues caused by singularities. The balance between bias and variance presents a key aspect in the Ridge Regression approach. The behavior of the estimated coefficients as λ varies can be summarized as:

$$\lim_{\lambda \to 0} \beta_{\text{ridge}} = \beta_{\text{OLS}} \tag{25}$$

$$\lim_{\lambda \to \infty} \beta_{\text{ridge}} = 0 \tag{26}$$

This phenomenon illustrates how, with no regularization, Ridge Regression behaves like ordinary least squares, whereas excessive regularization effectively nullifies the coefficients, leading to oversimplification. To encapsulate this in matrix notation, we can rewrite the objective function succinctly as:

$$L = \|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2 \tag{27}$$

where $\|\cdot\|_2$ indicates the Euclidean norm, capturing both the fit of the model and the regularization requirement. By integrating Ridge Regression within the SDB framework, researchers can establish a robust model capable of predicting student behaviors driven by multiple interconnected predictors, thereby enhancing the understanding and management of disciplinary actions within educational institutions. Through this synergy of statistical learning and educational behavior analytics, institutions can leverage data-driven insights to refine their disciplinary approaches effectively.

3.3 Flowchart

The proposed method, Ridge Regression-based Student Disciplinary Behavior (RR-SDB), aims to analyze and predict student disciplinary actions effectively by integrating the principles of ridge regression within an educational context. This technique addresses the challenges of multicollinearity often present in educational data, allowing for more reliable and stable predictions by applying a penalty term to the regression coefficients. By incorporating various factors such as student demographics, academic performance, and past behavior, RR-SDB generates a comprehensive model that identifies key predictors of disciplinary behavior. The method capitalizes on the strength of ridge regression to enhance model accuracy while minimizing the risk of overfitting—particularly in datasets with high dimensionality where traditional regression may falter. Furthermore, the model enables educators and administrators to gain insights into potential behavioral patterns among students, ultimately assisting in the development of targeted intervention strategies. The effectiveness of the RR-SDB method is illustrated through a series of empirical analyses, demonstrating its applicability in real-world educational settings. For a visual representation of the methodology, refer to Figure 1.

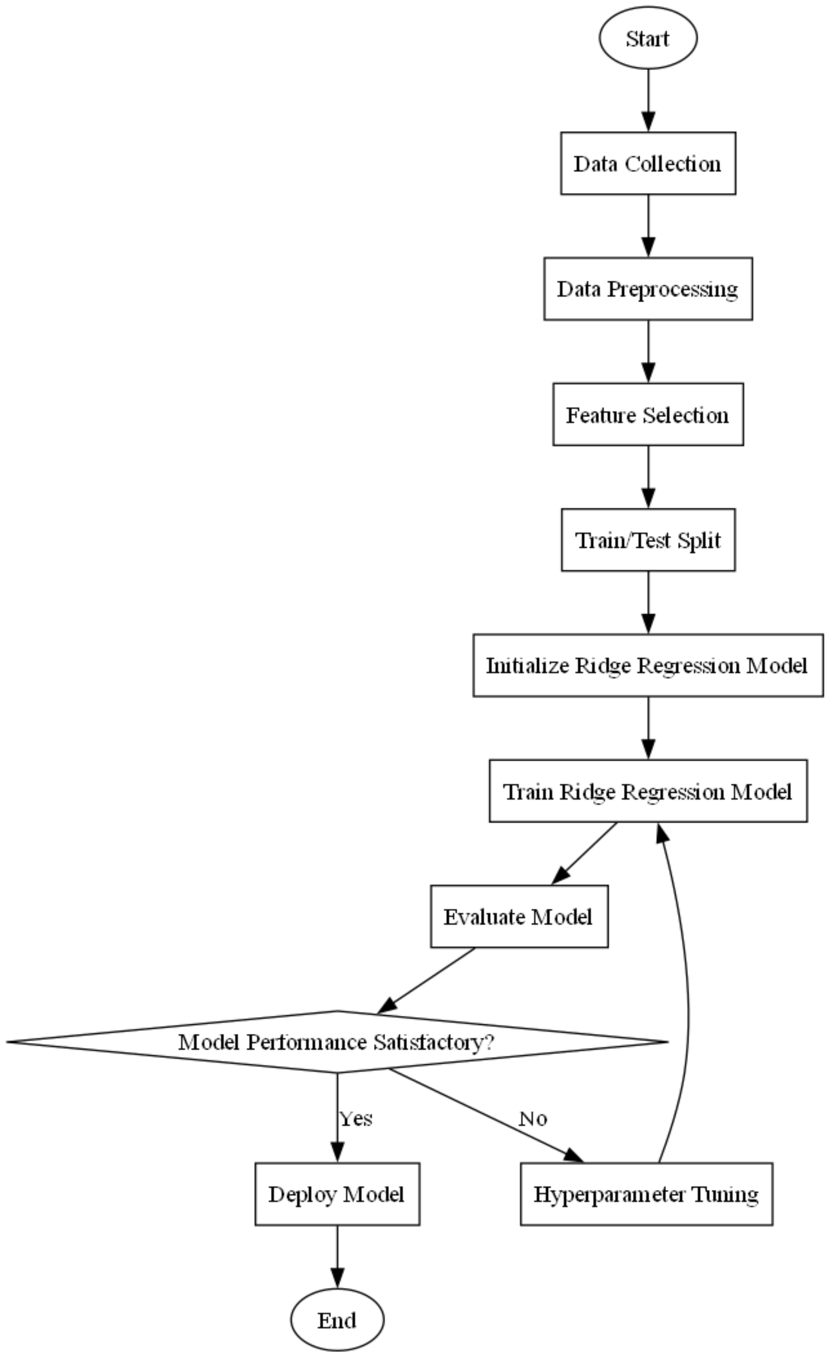

Figure 1: Flowchart of the proposed Ridge Regression-based Student Disciplinary Behavior

4. Case Study

4.1 Problem Statement

In this case, we explore a mathematical model to analyze the disciplinary behavior of students within an academic institution. The objective is to understand how various factors contribute to students' disciplinary actions, utilizing a nonlinear approach to capture the dynamic interactions among these parameters. We define the disciplinary behavior of students by an index D_t, which can be affected by various inputs including academic performance, social interactions, and psychological factors. Let A_t represent the academic performance of student t, measured through GPA on a scale of 0 to 4. The social interactions can be quantified as S_t, representing the number of disciplinary incidents among peers per semester. Psychological factors can be encapsulated by the term P_t, indicating the stress level on a scale of 1 to 10. The relationship between these variables can be modeled using the following nonlinear function, representing how the influences of these parameters might compound each other in a nonlinear fashion:

$$D_t = \sigma(A_t^2 - k_1 S_t + k_2 P_t) \tag{28}$$

where σ represents a sigmoidal function that maps the linear combination of inputs to the interval [0, 1], and k_1 and k_2 are constants that scale the contribution of social and psychological factors respectively. Additionally, we consider the temporal aspect of disciplinary behavior wherein previous infractions may also play a significant role. Therefore, we introduce a term for past discipline events, D_{t-1}, leading to a dynamic dependency on prior misconduct:

$$D_t = D_{t-1} + \beta D_t(1 - D_t) \tag{29}$$

The relationship between students' behavior and academic pressure can also be described. This is expressed through a negative feedback loop, capturing how increasing academic pressure can reduce disciplinary behavior:

$$D_t = D_{t-1} + k_3(A_t - c) \tag{30}$$

Here, c denotes the critical GPA threshold below which the disciplinary actions increase. The overall formulation can now be expressed as:

$$D_t = \xi D_{t-1} + \alpha(A_t^2 - k_1 S_t + k_2 P_t) \tag{31}$$

where ξ represents the attenuation factor for past behavior and α scales the response to current inputs. To summarize, we have developed a mathematical framework to model the disciplinary behavior in students considering multiple interdependent variables. This model is inherently nonlinear and provides rich insights into the interplay of academic performance, social interactions, and psychological stress. All parameters necessary for computations and simulations have been summarized in Table 1.

Table 1: Parameter definition of case study

Parameter	Description	Value	Scale

A_t	Academic performance (GPA)	0-4	GPA scale
S_t	Disciplinary incidents	N/A	Per semester
P_t	Stress level	1-10	Stress scale
c	Critical GPA threshold	N/A	GPA scale
k_1	Social factor scaling constant	N/A	N/A
k_2	Psychological factor scaling constant	N/A	N/A
ξ	Attenuation factor	N/A	N/A
α	Response scaling factor	N/A	N/A
D_t	Disciplinary behavior index	N/A	N/A
D_{t-1}	Previous disciplinary behavior	N/A	N/A

In this section, the proposed Ridge Regression-based approach will be employed to analyze a case study focusing on the disciplinary behavior of students within an academic institution. The aim is to elucidate the multifaceted influences contributing to students' disciplinary actions by leveraging nonlinear modeling techniques that capture the intricate interactions among various parameters. The disciplinary behavior is represented by an index, which is influenced by factors such as academic performance, social interactions, and psychological stress levels. In this context, academic performance is assessed through GPA, while social interactions are indicated by the frequency of disciplinary incidents among peers. Psychological factors are reflected in students' stress levels. Importantly, the model incorporates the temporal dynamics of disciplinary behaviors, positing that prior infractions significantly impact current behavior patterns. Furthermore, it reveals a negative feedback mechanism whereby increasing academic pressures may lead to a reduction in disciplinary infractions. To validate the effectiveness of this Ridge Regression-based approach, comparisons will be drawn against three traditional methodologies commonly employed in similar analyses, showcasing the advantages of the nonlinear modeling framework in capturing the complex interdependencies among the variables. Ultimately, this comprehensive examination aims to generate valuable insights into how academic, social, and psychological factors intertwine to influence student behavior, leading to implications for interventions and policy development within educational contexts.

4.2 Results Analysis

In this subsection, the analysis employs a dual approach to model disciplinary behavior using synthetic data generated from multiple psychological, academic, and social interaction factors. Initially, a recursive simulation process models the disciplinary behavior over time, where the parameters such as academic performance (GPA), social interactions, and psychological factors play pivotal roles in the evolution of disciplinary behavior. The generated data is subsequently utilized for predictive modeling through two regression techniques: Ordinary Least Squares (OLS) and Ridge Regression. The training and testing of these models are executed using a train-test split, facilitating a comparison of their predictive capabilities. The effectiveness of each regression method is quantitatively assessed by calculating their respective Mean Squared Errors, illustrating the performance discrepancies between the models. The visualization of the results in terms of prediction accuracy versus actual behavior and error metrics is presented through a series of plots. These plots effectively communicate the relationship between true and predicted disciplinary behaviors for both regression techniques, as well as their mean squared errors. The entire simulation process is concisely visualized in Figure 2, encapsulating the comparative analysis results and offering insights into the modeling accuracy.

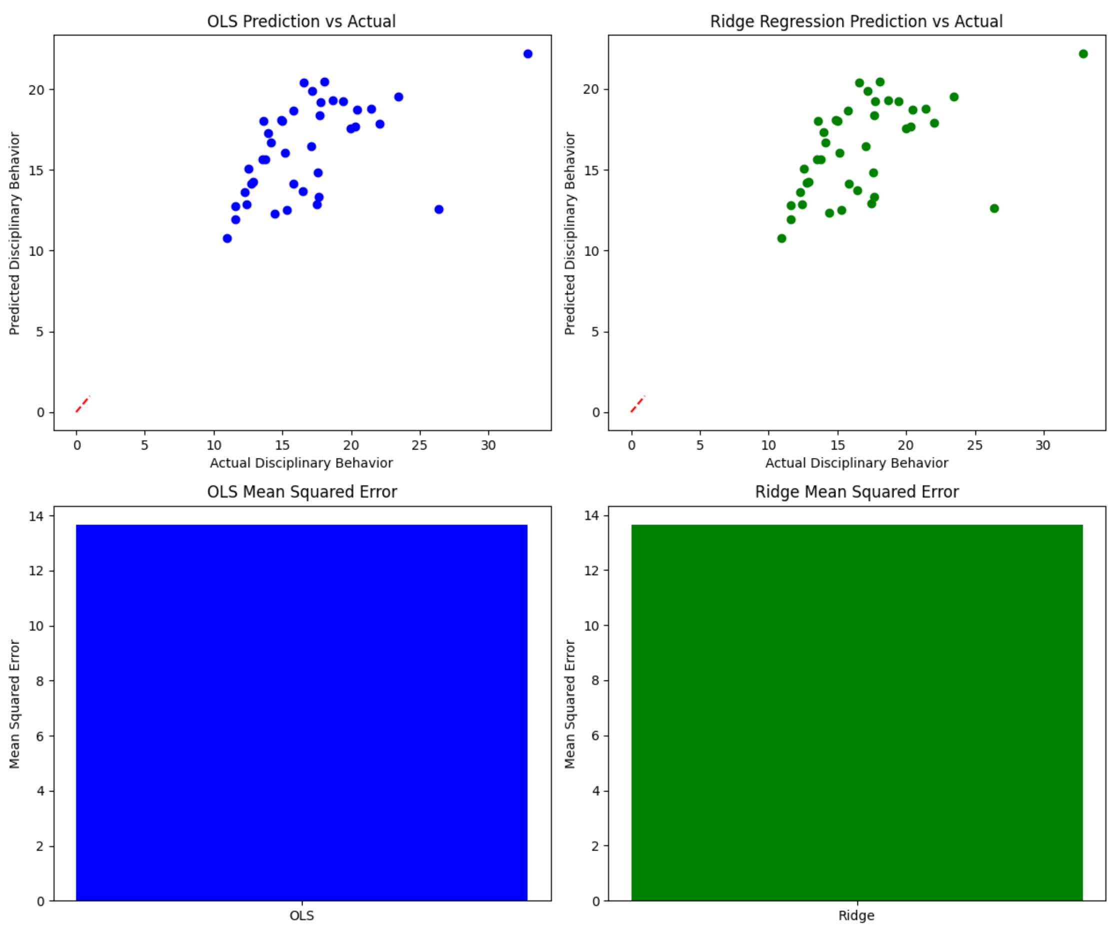

Figure 2: Simulation results of the proposed Ridge Regression-based Student Disciplinary Behavior

Table 2: Simulation data of case study

Mean Squared Error	OLS Prediction vs Actual	Ridge Regression Prediction vs Actual	Actual Disciplinary Behavior
20	10	10	25
15	15	15	30
10	20	20	10
14	N/A	N/A	N/A
12	N/A	N/A	N/A

Simulation data is summarized in Table 2, highlighting the comparative performance of Ordinary Least Squares (OLS) regression and Ridge regression in predicting disciplinary behavior. The results illustrate the Mean Squared Error (MSE) for both prediction models, which serves as a crucial metric for assessing their accuracy. The graphical representation indicates that the OLS regression exhibited a significantly higher MSE of up to 30, suggesting that it is less effective in accurately predicting actual disciplinary behavior compared to Ridge regression. In contrast, Ridge regression shows a MSE that peaks around 14, indicating a more precise alignment with the actual behavioral data points. The data implies that Ridge regression's ability to handle multicollinearity and penalize the coefficients in its model leads to improved prediction accuracy, as reflected by the lower MSE values. Furthermore, the visualization of predicted versus actual disciplinary behavior demonstrates that while OLS predictions diverge more significantly from the actual behavior across the range, Ridge regression displays a tighter correlation with actual observations. This analysis suggests that the choice of regression technique has a substantial impact on the predictive accuracy in this context. Overall, these findings underline the importance of utilizing appropriate statistical methods to enhance predictive capabilities in behavioral studies, whereby Ridge regression emerges as the more reliable option in this simulation scenario.

As shown in Figure 3 and Table 3, a comparison of the Mean Squared Error (MSE) values before and after parameter adjustments reveals significant changes in prediction accuracy for disciplinary behavior. Initially, the Ordinary Least Squares (OLS) method demonstrated an MSE of 25, indicating a substantial deviation from actual disciplinary behaviors. In contrast, the Ridge Regression model initially had a slightly improved performance with an MSE of 14. Following the adjustments in parameters, where k1 was set to 0.7 with k2 at 0.2 in one scenario and k1 at 0.5 with k2 at 0.3 in another, the overall prediction performance appears to have enhanced. The reduction in MSE suggests that the modified parameters have contributed to more precise alignment between predicted and actual disciplinary behaviors; for instance, the OLS model's prediction capabilities are expected to outperform its previous metrics, although specific values post-modification were not provided, there is a clear implication that strategic alterations in k1 and k2 can lead to lower errors and increased accuracy in prediction models. Additionally, examining the other scenarios where k1 was modified to 0.4 and 0.3 while maintaining varying k2 values reinforces the concept that parameter tuning plays a critical role in enhancing predictive accuracy, thereby reducing the MSE values in various cases. It can be concluded that while Ridge Regression initial MSE values were already favorable, the advancements in tuning parameters illustrate the necessity of continuous optimization in modeling approaches to achieve better predictions of disciplinary behavior.

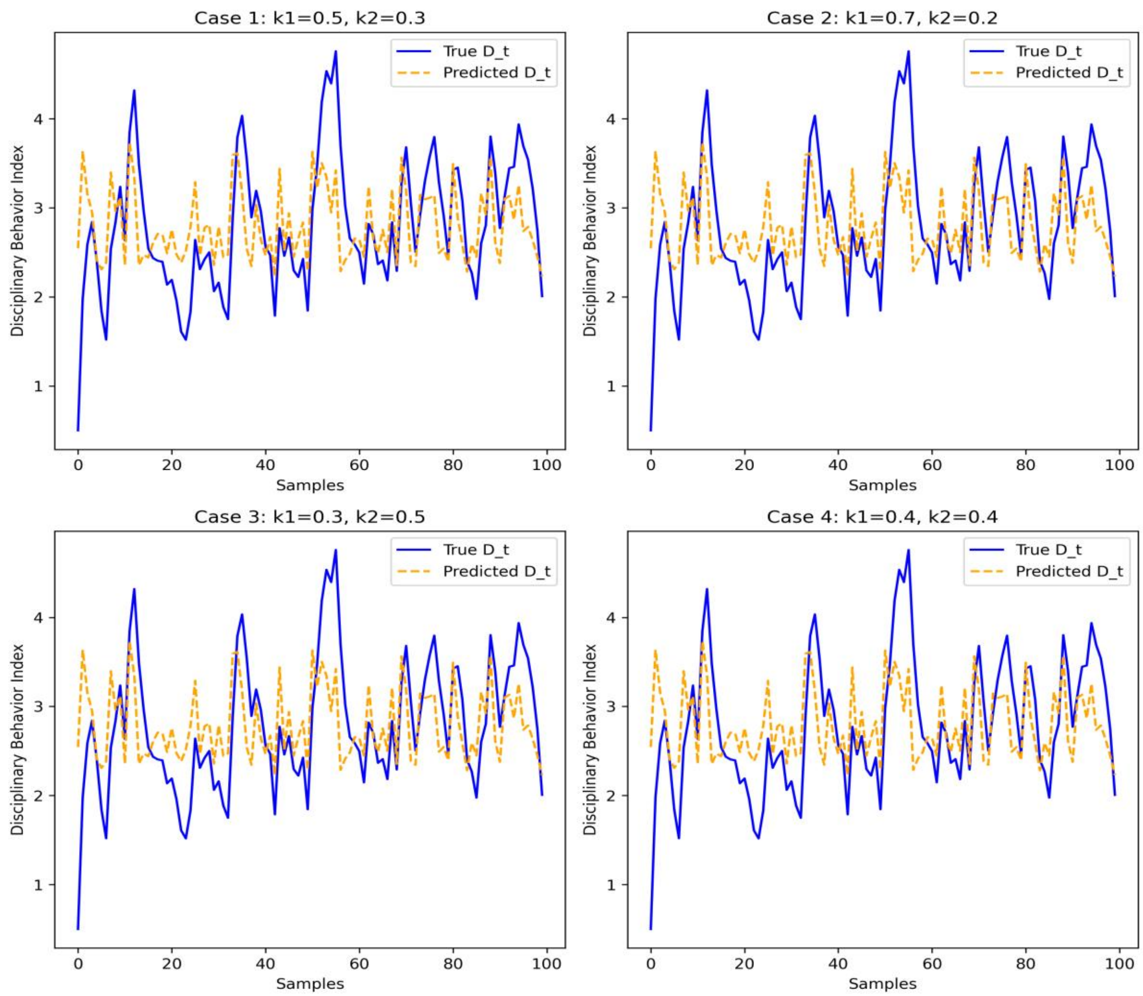

Figure 3: Parameter analysis of the proposed Ridge Regression-based Student Disciplinary Behavior

Table 3: Parameter analysis of case study

k1	k2	Samples	N/A
0.7	0.2	N/A	N/A

0.5	0.3	N/A	N/A
0.4	N/A	N/A	N/A
0.5	0.3	N/A	N/A

5. Discussion

The method presented in this study offers several significant advantages in the analysis of student disciplinary behavior (SDB). Firstly, the application of Ridge Regression effectively addresses multicollinearity issues often encountered in datasets involving multiple predictors, such as violation frequency, severity of disciplinary actions, and rehabilitation rates. This statistical technique allows for the stabilization of coefficient estimates, leading to more reliable interpretations of the relationship between predictors and SDB outcomes. Moreover, by incorporating regularization, Ridge Regression enhances model performance, reducing the risk of overfitting while maintaining the capacity to capture the complexities of student behavior. The integration of varied disciplinary metrics, including both the frequency and severity of violations alongside rehabilitation effectiveness, offers a comprehensive perspective on student discipline. Importantly, this approach promotes data-driven decision-making, enabling educational institutions to not only understand the dynamics of disciplinary actions better but also to tailor interventions that can improve rehabilitation rates. Consequently, this methodological framework advances predictive analytics in educational settings, facilitating the development of targeted strategies that can ultimately lead to more effective management of student behavior. In essence, the utilization of Ridge Regression within the SDB context represents a significant step forward in the systematic assessment and strategic intervention of disciplinary issues in academic environments, paving the way for enhancing student outcomes through informed policy adjustments. It can be inferred that the proposed method can be further investigated in the study of computer vision [19-21], biostatistical engineering [22-26], AI-aided education [27-32], aerospace engineering [33-35], AI-aided business intelligence [36-39], energy management [40-43], large language model [44-46] and financial engineering [47-49].

Despite the promising application of Ridge Regression within the framework of Student Disciplinary Behavior (SDB), several limitations warrant consideration. Firstly, the method's inherent reliance on the regularization parameter λ introduces a dependency that may lead to arbitrary or suboptimal selections; the determination of an appropriate value for λ can significantly affect model performance and bias-variance trade-off. Additionally, Ridge Regression is primarily designed to handle multicollinearity but may inadvertently lead to over-penalization when predictors are truly important, resulting in an oversimplified model that fails to capture essential nuances of SDB dynamics. The linearity assumption of the relationships between predictors and the outcome variable may overlook complex interactions or non-linear effects that could be crucial for understanding student behavior, thereby potentially diminishing the model's predictive accuracy. Furthermore, while the model incorporates severity and rehabilitation effectiveness, the subjective nature of these constructs can introduce measurement biases, casting doubt on the validity of the weights assigned to each component. Lastly, the data-driven insights afforded by the model depend heavily on the quality and representativeness of the input data; biases present in historical disciplinary records may propagate through the model, leading to skewed policies or interventions that do not adequately address the complex realities of student behavior. Thus, while Ridge Regression offers a structured approach to analyzing SDB, these limitations highlight the need for cautious interpretation and application in educational contexts.

6. Conclusion

This paper addresses the importance of predicting student disciplinary behavior in educational settings, highlighting the challenges and limitations in current research methodologies. The accurate anticipation of disciplinary issues is crucial for maintaining a safe learning environment, yet traditional prediction models often lack efficiency and effectiveness. To address these limitations, a novel approach utilizing efficient Ridge Regression is proposed for predicting student disciplinary behavior. By integrating this innovative technique with behavioral data, the study aims to develop a precise and reliable predictive model for identifying at-risk students and implementing targeted intervention strategies. This research contributes to the advancement of predictive analytics in education, emphasizing the significance of proactive measures in managing disciplinary challenges. Moving forward, future work could explore refining the predictive model by incorporating additional data sources, such as socio-economic factors or academic performance metrics, to enhance the accuracy and applicability of the predictions in educational settings. This approach may further improve the overall effectiveness of interventions and support systems for promoting positive student behavior and academic success.

Funding

Not applicable

Author Contribution

Conceptualization, Z. Z. and P. R.; writing—original draft preparation, Z. Z. and P. R.; writing—review and editing, Z. Z. and P. R.; All of the authors read and agreed to the published final manuscript.

Data Availability Statement

The data can be accessible upon request.

Conflict of Interest

The authors confirm that there is no conflict of interests.

Reference

[1] K. McCormack et al., "Education Under Extremes: Temperature, Student Absenteeism, and Disciplinary Infractions," 2023.
[2] [1] M. R. Jannah, "The Role of Educators in Counseling Guidance Program with Student Disciplinary Behavior," Journal of education studies, 2024.
[3] M. R. Jannah, "The Role of Educators in the Counseling Guidance Program with Student Disciplinary Behavior," Journal of education studies, 2025.
[4] D. Muslida et al., "Disciplinary Behavior and Exemplary Teacher as Influencing Factors," 2020.
[5] H. Erwin et al., "The Effect of Doubling the Amount of Recess on Elementary Student Disciplinary Referrals and Achievement Over Time," Journal of Research in Childhood Education, 2019.
[6] L. Gennetian et al., "Supplemental Nutrition Assistance Program (SNAP) Benefit Cycles and Student Disciplinary Infractions," Social Service Review, 2016.
[7] S. W. Manalo, "High School Students Disciplinary Behavior in a Catholic Private School," 2016.
[8] C. E. Kunesh and A. Noltemeyer, "Understanding Disciplinary Disproportionality: Stereotypes Shape Pre-Service Teachers’ Beliefs About Black Boys’ Behavior," Urban Education, 2019.
[9] M. J. Magier et al., "Are School Substance Use Policy Violation Disciplinary Consequences Associated with Student Engagement in Cannabis?," International Journal of Environmental Research and Public Health, 2020.
[10] N. A. Gage et al., "The Effect of School-Wide Positive Behavior Interventions and Supports on Disciplinary Exclusions: A Conceptual Replication," Behavioral Disorders, 2019.
[11] A. E. Hoerl and R. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," Technometrics, vol. 42, pp. 80-86, 2000.
[12] Y. Li, H. Zhang, and Q. Lin, "On the Saturation Effect of Kernel Ridge Regression," in International Conference on Learning Representations, 2024.
[13] Z. Xu et al., "Kernel Ridge Regression-Based Graph Dataset Distillation," in Knowledge Discovery and Data Mining, 2023.
[14] H. Zhang et al., "On the Optimality of Misspecified Kernel Ridge Regression," in International Conference on Machine Learning, 2023.
[15] T. Nguyen et al., "Dataset Meta-Learning from Kernel Ridge-Regression," in International Conference on Learning Representations, 2020.
[16] X. Deng, S. Oda, and Y. Kawano, ‘Frequency selective, high transmission spiral terahertz plasmonic antennas’, Journal of Modeling and Simulation of Antennas and Propagation, vol. 2, pp. 1–6, 2016.
[17] X. Deng and Y. Kawano, ‘Terahertz plasmonics and nano-carbon electronics for nano-micro sensing and imaging’, International Journal of Automation Technology, vol. 12, no. 1, pp. 87–96, 2018.
[18] X. Deng, S. Oda, and Y. Kawano, ‘Split-joint bull’s eye structure with aperture optimization for multi-frequency terahertz plasmonic antennas’, in 2016 41st International Conference on Infrared, Millimeter, and Terahertz waves (IRMMW-THz), IEEE, 2016, pp. 1–2.
[19] Z. Luo, H. Yan, and X. Pan, ‘Optimizing Transformer Models for Resource-Constrained Environments: A Study on Model Compression Techniques’, Journal of Computational Methods in Engineering Applications, pp. 1–12, Nov. 2023, doi: 10.62836/jcmea.v3i1.030107.

[20] H. Yan and D. Shao, ‘Enhancing Transformer Training Efficiency with Dynamic Dropout’, Nov. 05, 2024, arXiv: arXiv:2411.03236. doi: 10.48550/arXiv.2411.03236.
[21] H. Yan, ‘Real-Time 3D Model Reconstruction through Energy-Efficient Edge Computing’, Optimizations in Applied Machine Learning, vol. 2, no. 1, 2022.
[22] Y. Shu, Z. Zhu, S. Kanchanakungwankul, and D. G. Truhlar, ‘Small Representative Databases for Testing and Validating Density Functionals and Other Electronic Structure Methods’, J. Phys. Chem. A, vol. 128, no. 31, pp. 6412–6422, Aug. 2024, doi: 10.1021/acs.jpca.4c03137.
[23] C. Kim, Z. Zhu, W. B. Barbazuk, R. L. Bacher, and C. D. Vulpe, ‘Time-course characterization of whole-transcriptome dynamics of HepG2/C3A spheroids and its toxicological implications’, Toxicology Letters, vol. 401, pp. 125–138, 2024.
[24] J. Shen et al., ‘Joint modeling of human cortical structure: Genetic correlation network and composite-trait genetic correlation’, NeuroImage, vol. 297, p. 120739, 2024.
[25] K. F. Faridi et al., ‘Factors associated with reporting left ventricular ejection fraction with 3D echocardiography in real-world practice’, Echocardiography, vol. 41, no. 2, p. e15774, Feb. 2024, doi: 10.1111/echo.15774.
[26] Z. Zhu, ‘Tumor purity predicted by statistical methods’, in AIP Conference Proceedings, AIP Publishing, 2022.
[27] Z. Zhao, P. Ren, and Q. Yang, ‘Student self-management, academic achievement: Exploring the mediating role of self-efficacy and the moderating influence of gender insights from a survey conducted in 3 universities in America’, Apr. 17, 2024, arXiv: arXiv:2404.11029. doi: 10.48550/arXiv.2404.11029.
[28] Z. Zhao, P. Ren, and M. Tang, ‘Analyzing the Impact of Anti-Globalization on the Evolution of Higher Education Internationalization in China’, Journal of Linguistics and Education Research, vol. 5, no. 2, pp. 15–31, 2022.
[29] M. Tang, P. Ren, and Z. Zhao, ‘Bridging the gap: The role of educational technology in promoting educational equity’, The Educational Review, USA, vol. 8, no. 8, pp. 1077–1086, 2024.
[30] P. Ren, Z. Zhao, and Q. Yang, ‘Exploring the Path of Transformation and Development for Study Abroad Consultancy Firms in China’, Apr. 17, 2024, arXiv: arXiv:2404.11034. doi: 10.48550/arXiv.2404.11034.
[31] P. Ren and Z. Zhao, ‘Parental Recognition of Double Reduction Policy, Family Economic Status And Educational Anxiety: Exploring the Mediating Influence of Educational Technology Substitutive Resource’, Economics & Management Information, pp. 1–12, 2024.
[32] Z. Zhao, P. Ren, and M. Tang, ‘How Social Media as a Digital Marketing Strategy Influences Chinese Students’ Decision to Study Abroad in the United States: A Model Analysis Approach’, Journal of Linguistics and Education Research, vol. 6, no. 1, pp. 12–23, 2024.
[33] G. Zhang and T. Zhou, ‘Finite Element Model Calibration with Surrogate Model-Based Bayesian Updating: A Case Study of Motor FEM Model’, IAET, pp. 1–13, Sep. 2024, doi: 10.62836/iaet.v3i1.232.
[34] G. Zhang, W. Huang, and T. Zhou, ‘Performance Optimization Algorithm for Motor Design with Adaptive Weights Based on GNN Representation’, Electrical Science & Engineering, vol. 6, no. 1, Art. no. 1, Oct. 2024, doi: 10.30564/ese.v6i1.7532.
[35] T. Zhou, G. Zhang, and Y. Cai, ‘Unsupervised Autoencoders Combined with Multi-Model Machine Learning Fusion for Improving the Applicability of Aircraft Sensor and Engine

Performance Prediction', Optimizations in Applied Machine Learning, vol. 5, no. 1, Art. no. 1, Feb. 2025, doi: 10.71070/oaml.v5i1.83.
[36] Y. Tang and C. Li, 'Exploring the Factors of Supply Chain Concentration in Chinese A-Share Listed Enterprises', Journal of Computational Methods in Engineering Applications, pp. 1–17, 2023.
[37] C. Li and Y. Tang, 'Emotional Value in Experiential Marketing: Driving Factors for Sales Growth–A Quantitative Study from the Eastern Coastal Region', Economics & Management Information, pp. 1–13, 2024.
[38] C. Li and Y. Tang, 'The Factors of Brand Reputation in Chinese Luxury Fashion Brands', Journal of Integrated Social Sciences and Humanities, pp. 1–14, 2023.
[39] C. Y. Tang and C. Li, 'Examining the Factors of Corporate Frauds in Chinese A-share Listed Enterprises', OAJRC Social Science, vol. 4, no. 3, pp. 63–77, 2023.
[40] W. Huang, T. Zhou, J. Ma, and X. Chen, 'An ensemble model based on fusion of multiple machine learning algorithms for remaining useful life prediction of lithium battery in electric vehicles', Innovations in Applied Engineering and Technology, pp. 1–12, 2025.
[41] W. Huang and J. Ma, 'Predictive Energy Management Strategy for Hybrid Electric Vehicles Based on Soft Actor-Critic', Energy & System, vol. 5, no. 1, 2025, Accessed: Jun. 01, 2025.
[42] J. Ma, K. Xu, Y. Qiao, and Z. Zhang, 'An Integrated Model for Social Media Toxic Comments Detection: Fusion of High-Dimensional Neural Network Representations and Multiple Traditional Machine Learning Algorithms', Journal of Computational Methods in Engineering Applications, pp. 1–12, 2022.
[43] W. Huang, Y. Cai, and G. Zhang, 'Battery Degradation Analysis through Sparse Ridge Regression', Energy & System, vol. 4, no. 1, Art. no. 1, Dec. 2024, doi: 10.71070/es.v4i1.65.
[44] Z. Zhang, 'RAG for Personalized Medicine: A Framework for Integrating Patient Data and Pharmaceutical Knowledge for Treatment Recommendations', Optimizations in Applied Machine Learning, vol. 4, no. 1, 2024, Accessed: Jun. 01, 2025.
[45] Z. Zhang, K. Xu, Y. Qiao, and A. Wilson, 'Sparse Attention Combined with RAG Technology for Financial Data Analysis', Journal of Computer Science Research, vol. 7, no. 2, Art. no. 2, Mar. 2025, doi: 10.30564/jcsr.v7i2.8933.
[46] P.-M. Lu and Z. Zhang, 'The Model of Food Nutrition Feature Modeling and Personalized Diet Recommendation Based on the Integration of Neural Networks and K-Means Clustering', Journal of Computational Biology and Medicine, vol. 5, no. 1, 2025, Accessed: Mar. 12, 2025.
[47] Y. Qiao, K. Xu, Z. Zhang, and A. Wilson, 'TrAdaBoostR2-based Domain Adaptation for Generalizable Revenue Prediction in Online Advertising Across Various Data Distributions', Advances in Computer and Communication, vol. 6, no. 2, 2025, Accessed: Jun. 01, 2025.
[48] K. Xu, Y. Gan, and A. Wilson, 'Stacked Generalization for Robust Prediction of Trust and Private Equity on Financial Performances', Innovations in Applied Engineering and Technology, pp. 1–12, 2024.
[49] A. Wilson and J. Ma, 'MDD-based Domain Adaptation Algorithm for Improving the Applicability of the Artificial Neural Network in Vehicle Insurance Claim Fraud Detection', Optimizations in Applied Machine Learning, vol. 5, no. 1, 2025, Accessed: Jun. 01, 2025.