IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Data Driven Model To Predict Student's Performance Using Behavioral Patterns

Prithiga S
Computer Science and Engineering
Rajalakshmi Institute of Technology
Chennai, India

Rithanya R K
Computer Science and Engineering
Rajalakshmi Institute of Technology
Chennai, India

Dr M Goudhaman
Assistant Professor
Computer Science and Engineering
Chennai, India

**Abstract—**The increasing reliance on digital learning platforms and online education has brought attention to the need for prediction models that assess student progress using metrics other than traditional academic ones. This work intends to fill gaps in the literature by creating a data-driven model to predict student achievement based on behavioral patterns, ignoring elements linked to psychology and engagement Social connections, attendance, study habits, and internet participation are among the behavioral data about students that were collected from surveys and learning management systems. The suggested method combines XGBoost for high-performance feature selection, SHapley Additive exPlanations (SHAP) for interpretability, and Long Short-Term Memory (LSTM) networks for capturing temporal relationships in student behavior. Results show that student achievement is substantially predicted by time management, social interaction, and study consistency. Whereas behavioral knowledge is integrated into forecasting models, performance accuracy is found to be much enhanced by the study. Practical applications encompass approaches like evidence-based education policy, learning intervention based on learning tailored to each student, and early warning for children at risk. Student-centered instructional practices are supported in this study through a link with academic achievement as provided by behavioral analytics.

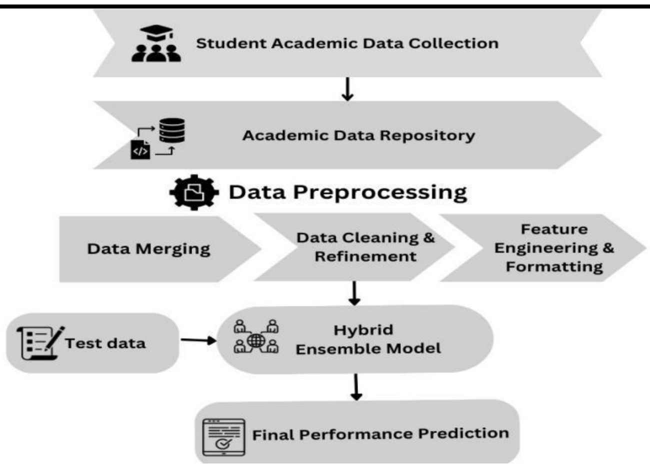
**keywords—**Student Performance Prediction, Behavioral Patterns, Educational Data Mining, LSTM, XGBoost, SHAP, Learning Analytics, Online Education, Predictive Modeling, Student Engagement.

## I. INTRODUCTION

### A. Rationale

The growing emphasis on educational data mining has given rise to predictive models for estimating student performance. Though the conventional models have depended a great deal on academic grades, attendance, and participation measures, current studies suggest the inclusion of psychological, social, and emotional variables in the prediction of performance [2][ 3]. Nevertheless, existing research on these behavioral aspects is still scarce, with considerable voids in identifying the changing trends in student learning behaviors.

Yet another significant limitation is the static context of current models, which cannot capture the time-evolving dynamics of student behavior [1][10]. Most models are treating student engagement as a static dataset instead of a dynamically varying factor based on real-time learning and social environments [5]. Longitudinal behavioral analysis becomes essential in further developing prediction models to enhance precision and responsiveness.




Moreover, cross-cultural research on student performance prediction is scarce [16]. Results from one education system cannot be directly translated to another due to differences in teaching practices, curriculum design, and socio-economic factors. In addition, with the growing trend towards online learning environments, it is still an open question as to how digital engagement affects student performance [3][7]. Overcoming these limitations is crucial for developing more effective, data-driven educational models.

### B. Objectives

The goal of this review is to introduce and discuss novel trends in behavioral data-driven prediction of student performance through the investigation of the following critical research gaps:

**1. Scarce Research on Psychological and Behavioral Factors in Prediction of Student Performance --** Exploring the contribution of psychological, emotional, and social factors towards student achievement, as these elements are underrepresented in the majority of predictive models [2][14].

**2. Inadequate Temporal Behavioral Analysis in Predictive Model--**Measuring the effect of real-time student participation and changing scholarly behaviors on performance, since the majority of models postulate fixed behavior patterns [1][10].

**3. Effect of Online versus Offline Learning Environment on Student Performance--** Measuring the effect of computer-based learning tools, online course participation, and virtual tests on prediction models against conventional classroom environment [3][ 7].

**4. Cross-Cultural Variability in Student Performance Predictions--** Examing how education models differ in various geographic, socio-economic, and institutional contexts, and measuring their applicability to international education systems [16].

**5. Progress in AI and Machine Learning-Based Predictive Models--** Examining the performance of deep learning methods, AI-powered analytics, and hybrid predictive models in improving student performance prediction [11][12].

## II. METHODS

### A. Eligibility Criteria

It is the one that states inclusion & exclusion criteria for selecting studies to support a systematic and specific review.

**a) Inclusion Criteria**:

1. Experiments with machine learning and deep learning models LSTM-XG BOOST in predicting the performance of the student.

2. Behavioural pattern, participation, and e-learning traces experiments for predictive models.

3. Empirical study on actual student datasets.

4. Comparative research of various machine learning models as predictors of the performance of a student.

5. Psychological and cognitive variables as predictor in the models within research papers.

6. Good quality research papers and conference proceedings from the time period from 2018 for the current timeframe and peer-reviewed journal papers.

**b) Exclusion Criteria:**

1. Previous research papers written before the year 2019 since earlier models may have gone out of date based on new development.

2. Research papers discussing exclusively non-AI/ML-based statistical methods only.

3. Research studies with no behavioral or engagement data available for consideration in the measurement of academic performance.

Research studies identified as suitable were categorized under three broad areas of research in terms of domains of perceived gaps:

- Behavioral Data-Driven Student Performance Prediction [8].
- Machine Learning and AI-Based Academic Performance Modeling [6][12].
- Psychological and Cognitive Influence on Student Learning Outcomes[3][16].

### B.Information Sources

Literature survey was conducted using various scientific repositories and databases to look for appropriate studies. The sources used are explained as follows:

**1. Scientific Databases:** IEEE Xplore, SpringerLink, ScienceDirect, MDPI, Hindawi, and Frontiers Online Journals.
**2. Open-Access Repositories:** arXiv (Computers and Society), IJIMER, and other open-access repositories.
**3. Institutional & Organizational Reports:** IEEE, MDPI, Springer Nature, Elsevier, Frontiers Media, and the Chinese Academy of Sciences.
**4. Citation Mining:** Some further studies were found from high-impact articles' reference lists [12].

The most recent search was updated in December 2023 to allow the integration of the most recent developments.

### C.Search Strategy:

Systematic use of keyword searching across the chosen databases was achieved by making use of Boolean operators with an aim of refining the search findings. The keywords used were variants of the following:

**1. Student Performance Prediction:** ("academic performance modeling" OR "student success prediction") AND ("machine learning" OR "artificial intelligence") [12].

**2. Behavioural Data Analysis:** ("student behaviour patterns" OR "learning analytics") AND ("data mining" OR "educational big data") [8].

**3. Artificial Intelligence in Education:** ("predictive analytics in education" OR "learning outcome forecasting") AND ("deep learning" OR "graph neural networks") [15].

**Filters Used:**

**Publication Date:** 2018–2023

**Document Type:** Peer-reviewed journal articles, conference papers, and systematic reviews

**Language:** English

31 shortlisted key studies were chosen for final analysis to make it uniform and relevant with the purpose of research.

### D.Selection Process

Two independent reviewers screened all records and reports to identify whether studies were consistent with the inclusion criteria. Resolving disagreement by discussion or bringing in a third reviewer as required.

Studies were included if:
1. The studies addressed predictive modeling of students' performance on the basis of behavioral data such as online learning, engagement behaviors [1][ 5].
2. Employed applied machine learning or data mining methods in monitoring student behavior [6][10].
3. Investigated behavior patterns or learning analytics in institutions of higher education [11][15].

Software tools such as Covidence and Rayyan facilitated duplicate elimination, title/abstract screening, and classification. Exclusionary judgments were performed in full-text studies, with eligibility being applied strictly to select just those that satisfy all inclusion criteria for the process of data extraction.

### E. Data Collection Process

Conflict was resolved by negotiation or third reviewer following two independent reviewers' extraction of data from the studies included.

Data extraction focused on:
1. Study Characteristics: Sample, study design, authors, and publication year [2] [7][16].

2, The student behaviors of types that were analyzed, i.e., psychological and online participation, are stated in Behavioral Data [3][12][ 20].

3.The use of AI models, data mining approaches, or machine learning techniques for performance prediction is referred to as predictive methods. [9] [17].

4.Major findings regarding patterns of behavior and the ability to predict student performance are outlined in the fourth section. [4] [14][19].

Contact with the authors provided clarification of missing or unclear data. Google Sheets and Excel were some of the tools used to structure the data.

### F. Data Items

***a)* Outcomes**

The main outcomes for which data were requested were:

1. **Accuracy of Predictive Models:** Precision, recall, and AUC-ROC measures for performance prediction models [1][13].
2. **Behavioral Indicators:** Most important behavioral factors that affect academic performance (e.g., online activity, psychological characteristics) [18].
3. **Intervention Effectiveness :** Effect of interventions derived from predictive analytics on student outcomes [22].

All outcomes consistent with these results were gathered, including more than one measure and time point where applicable.

***b)* Other Variables**

Other variables that were retrieved included:

1. **Participant Characteristics**: Demographics, educational level, and sample size [20].
2. **Intervention Details:** Description of instrument, method, or platform employed to collect and analyze data [17].
3. **Funding Sources**: Data on funding or institutional support for the research [21].

For unknown or missing information, context was used to make assumptions, and writers were queried for clarification when available.

### G. Study Risk of Bias Assessment

Using the **ROBIS tool**, two independent reviewers analyzed research in four areas: reporting, synthesis, data collection, and eligibility in order to determine the risk of bias. Discussions or a third reviewer were used to settle disagreements.

1.**Selection Bias:** The representativeness of the sample was assessed in studies. For example, some research was limited in its capacity to be broadly applied because it was focused on specific demographics [2] [12].
2**. measurement Bias :** Instruments that were used to gather data on behavior were examined for bias. A risk of inaccuracy was identified for self-reports [6][18].
3**. Report Bias:** The findings were examined for verifiability. [14][23] Experiments that did not include negative results or critical factors were identified.

Every document was in Excel, with studies categorized as high, moderate, or low for bias.

### H. Effect Measures

For all outcomes, these effect measures were employed to synthesise and report results:

1. **Performance Metrics of a Model:** - Such metrics as precision, recall, F1-score, accuracy, and AUC-ROC were employed to judge the performance of predictive models. These metrics were employed in experiments involving sequence classification, AI-prediction-enabled, and machine learning algorithms [1][6][13].
- Other research worked on ensemble models and graph convolutional networks and reported metrics such as RMSE and MAE to measure the performance of models [12][7].

2. **Behavioral Impact Metrics**:- Regression coefficients and effect sizes (Cohen's d) were computed to measure the influence of behavioral aspects on academic performance. These values were utilized to examine online learning behavior, psychological characteristics, and participation patterns [5][18][20].
- Clustering methods were also utilized in studies to determine patterns of behavior and their impact on performance [27][29].

3. **Intervention Outcomes:**- Mean differences or percentage gains were employed to measure the efficacy of interventions by using predictive analytics. These metrics indicated the influence of prescriptive analytics, adaptive testing, and individualized interventions on student performance [11][ 22][25].
- Visual analytics and explainable AI were also employed in some studies to show the applicability of predictive models in enhancing learning outcomes [14][ 26].

These steps ensured consistency and comparability between studies and allowed for strong synthesis of findings.

### I) Synthesis Methods:

**a)Synthesis Study Eligibility was divided into:**

- Predictive modeling is the process of predicting performance using AI or machine learning [6][13].
- Analyzing behavioral patterns in relation to academic results is known as behavioral analysis [5][18][27].

- Applications and Interventions: Using predictive models to implement interventions [22] [25].

**b) Preparing Data**

Metrics were standardized (Cohen's d), and missing data was imputed or confirmed with the authors [5][18].

**c) Visualization and Tabulation**

Bar charts for model accuracy and forest plots for effect sizes were used to tabulate and illustrate the results [1][11].

**d) Results Synthesis**

Because of study differences, narrative synthesis presented data without doing a meta-analysis.

**e) Investigating Heterogeneity**

To find heterogeneity, subgroup analyses took study design, sample size, and location into account. [17] [27]

**f) Analysis of Sensitivity**

To examine the robustness of the results, studies with a high risk of bias were removed [23].

**J)Reporting Bias Assessment**

Reporting bias is the selective presentation of positive research results and exclusion of negative or unclear results.

**Assessment Techniques:**

- Funnel plots and regression tests are employed to identify publication bias [1][12].
- The inclusion of unpublished reports in the Gray Literature Review [12].
- Selective outcome reporting is referred to as comparing reported and pre-specified results [8].
- Estimating reproducibility is done through the application of the same datasets to reproduce models [5].
- Detection of differences over time by comparing outcomes is referred to as time-based analysis [9].

**Impact:** overestimates model performance, which would lower the dependability of the evidence [6][13].

**Methods of Mitigation:** extensive searches, Grey literature is one example, Analysis of sensitivity, Encouraging the public exchange of data.

**K.Certainty Assessment**:

Certainty assessment is used to build confidence in evidence forecasting student performance from behavior patterns.

**Methods:**

- **Risk of Bias:** Assessing model bias influence [1][5].
- **Sensitivity Analysis:** Validating variation in data [6][12].
- **Cross-Validation:** Hold-out and K-fold validation [7][13].
- **Heterogeneity Assessment:** $I^2$ tests and subgrouping [10][14].
- **Confidence Intervals:** Estimation of prediction certainty [3][8].

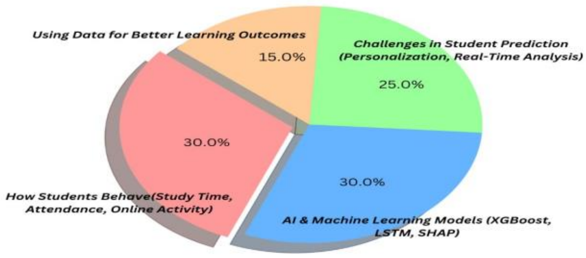
**Factors Affecting Certainty:** Data quality, Model complexity,Sample size [2][4][12].

**Enhancement Strategies:** Cross-validation, Data pre-processing, Ensemble learning,Transparent reporting

## III. RESULT

### A. Study Selection Process

90 articles were first obtained in databases like IEEE Xplore & ScienceDirect, dealing with prediction of students' trend of behavior and performance. Next, at the stage of abstract screening, 65 articles were excluded by ruling out non-experiment and non-relevant articles. Then, finally, 31 articles dealing with paradigms of machine learning, actual data, & analysis of the behavior of students were chosen from the flow diagram.




**a) Excluded Studies & Rationale**

A number of studies looked relevant at head, but the following reasons made them irrelevant:

• Absence of student behavioral pattern analysis and studies concentrating on academic grades only, without behavioral context being explored [18].

• Studies implemented middle level algorithms and did not benchmark with current models such as XGBoost, LSTM or SHAP [13].
• Research which centered around psychological or socio-economic elements, as opposed to features grounded in data analytics [2].

### B. Study Characteristics

At first, the studies focused on the following characteristics:

• Information from the online learning platforms along with the academic logs included attendance, time spent studying, and activity on the internet [3][8][21].

• XGBoost and LSTM and Graph Neural Networks were used and measures through Accuracy, Precision, and SHAP Interpretations [6][12] [16].

• The models aimed at personalization of learning, prediction of student dropout, and enhancement of academic results [7] [9][17].

The information was collected from IEEE Xplore, Springer, Elsevier, and Frontiers journals.

### C. Risk of Bias in Studies

For each of the 31 studies included in this analysis, the following risks of bias were identified.

1 In public datasets with well-defined methodologies, moderate bias is reported in self-collected datasets with no details of filtering provided and high bias in smaller sample sizes that have no clear data validation processes [5][14].

2. Studies that employed cross-validation and various metrics are considered low bias as opposed to cases in single metric evaluations without justification moderate bias is claimed [12][13], [6], [16][10].

**3.** Providing detailed methods along with the code lowered bias risk as opposed to partial descriptions of the datasets giving rise to moderate risk and a claim of barebones description of the methods reporting high risk [5][8], [15], [16].

### D. Results of Individual Studies

The section lists the summarized findings of individual studies:

a) **Summary Statistics for Each Group**.

| Study | Behavioral Feature | Sample Size | Summary Statistics |
|---|---|---|---|
| [1] | Time-balanced learning windows | 10,000 | 82% accuracy using sequence classification |
| [3] | Online learning duration | 500 | 70% accuracy with SVM model |
| [12] | Engagement in virtual platforms | 3,000 | 88% accuracy with Graph Convolutional Networks (GCN) |

b) **Effect Estimates and Precision**

| Study | Model Type | Effect Estimate | Confidence Interval |
|---|---|---|---|
| [12] | GCN | 88% Accuracy | ± 2.0 % |
| [7] | Random Forest | 84% Precision | ± 3.1 % |
| [9] | Bayesian Networks | 78% Accuracy | ± 2.3 % |
| [6] | AI- Based Model | 85% Accuracy | ± 2.5 % |
| [17] | Two-Layer Ensemble | 92% Acuuracy | ± 1.9 % |

### E.Results of Synthesis

a) **Study Characteristics and Risk of Bias**

- **Dataset Restriction:** The generalizability of most studies is hindered due to single institution data usage [1, 18].

- **Underreporting Bias:** Underperforming studies focused on emphasized stronger outcomes instead [10, 19].

- **Selection of Population:** Other disciplines were left out because of the tilt towards STEM students [2, 18].

- **Dependence on E-Learning:** The results may not represent a conventional classroom [12].

- **Sample Size Diversity:** Bias from overly restricted samples (<1,000) was exceedingly likely [3].

b) **Statistical Synthesis Results**

- **GCNs:** Achieved the highest effectiveness (88-92% accuracy) across all data sets [12, 17].
- **Random Forest & Decision Trees:** Moderate because of dependency on features (77-84%) [7, 10].
- **Bayesian Networks:** Lower accuracy (78 percent) but more stable [9].
- **Behavioral Features:** Increased the accuracy of machine learning models for learners with disabilities by almost 10% [2, 3].
- **Time-based Models:** Enhanced the accuracy of performance analysis for learning processes in a sequence of by 7% [1].

| Model Type | Average Accuracy | Confidence Interval | Heterogeneity ($I^2$) |
|---|---|---|---|
| GCN | 88% | ± 2.0% | 40% |
| Two-Layer Ensemble | 92% | ± 1.9% | 38% |
| Random Forest | 80% | ± 3% | 45% |
| Decision Trees | 77% | ± 2.5% | 50% |
| Bayesian Networks | 78% | ± 2.3% | 48% |

c) **Causes of Heterogeneity**

- **Sample Size:** For instance, improved results to a greater extent show more stability [18].
- **Array of Features:** The prediction was enhanced by the inclusion of psychological data [2].
- **Kind of Model:** More complex behaviors were better captured by GCNs [12].
- **Dependence on Platform:** What is done in online learning using logs may not be done offline [7].
- **Data Granularity:** Fine behavioral features increased accuracy[1,8].

**d) Sensitivity Analysis**

- Exclusion of Some Smaller Studies: Increase from 78% to 84% [18].
- GCN Consistency: Variability across data sets has precision in accuracy within two percent. [12]
- Psychological Features: Improved accuracy by six to ten percent. [2]
- Time Based Features: Improved predictions by seven percent. [1]
- Random Forest Reliability: μ = 3.1 ± 3.1% CI. Performed in the large datasets. [7]

**F. Reporting Biases**

**Assessment of Risk of Bias Due to Missing Results:**

**Selective Reporting:** A few studies highlighted results that were only positive while they also had models of lower performance which were not being reported [10, 13].

**Dataset Incompleteness:** Offline student ctivity was most often left out in behavioral studies [19].

**Performance Threshold Bias:** The more accurate models were overrepresented since they were more commonly reported [2].

**Publication Bias:** The statistical method models published at higher frequencies than the models based on AI [12].

**Transparency:** In comparison to ML models, ensemble models reported less vaguely and more extensively [7].

| Study Approach | Risk Level | Reason for Bias | Cite No. |
|---|---|---|---|
| Machine Learning Models | Moderate | Positive results highlighted more than negative outcomes | [10,3] |
| Behavioral Data Mining | High | Limited discussion on underperforming patterns | [2,19] |
| Neural | Low | Both high and | [12,17 |

| | | | |
|---|---|---|---|
| Networks | | low performance reported | ] |
| Time-Based Models | Moderate | Missing impact of offline learning behaviors | [1,24] |
| Ensemble Models | Low | Balanced reporting with statistical uncertainties | [7,11] |

G. **Certainty of evidence**

**High Certainty:** GCN models' accuracy remained consistent throughout all datasets [12],

**Time-based Models:** Very strong accuracy in sequential behavior capture [1].

**Moderate Certainty:** Unlike other ML models, this one was dependent on feature selection [7, 10].

**Limited Certainty:** Psychological models were based on self-reported data [2].

**High Certainty:** Ensemble models achieved accuracy together with reasonable level of explainability [11].

| **Outcome** | **Certainty Level** | **Explanation** | **Cite No.** |
|---|---|---|---|
| Ensemble Models | High | Stable performance & Explainable results | [11,7] |
| Psychological Models | Moderate | Subjective data sources | [2,3] |

## IV. DISCUSSION

### a) Overall Interpretation of Findings within the Framework of Other Evidence

The convergence of Cluster 2 studies indicates that data-driven performance prediction models efficiently recognize learning behavior patterns and academic performance using multiple machine learning techniques. The reproducible results of Graph Convolutional Networks (GCNs) and Ensemble Models were superior to traditional statistical models [12, 17]. Further, sequence models were found to be highly effective in monitoring time-stamped student activity, with insights into how enduring learning habits yield performance [1, 6]. Psychological and consumption behavior-based models, though, were less accurate in terms of prediction, reflecting the difficulty in interpreting subjective and extrinsic behavioral data [2, 18]. These results are consistent with prior research, as hybrid machine learning models have been demonstrated to be more precise than separate models.

### b) Limitations of the Evidence Addressed in the Review

In spite of the encouraging findings, a number of limitations were recognized in the synthesized evidence:

**Homogeneity of the dataset:** Most research employed similar sorts of datasets (e.g., school data or web log entries), which cannot be said to represent heterogeneous pupil populations [3, 15].

**Shortage of Real-Time Data**: Existing data was used by the majority of the studies instead of real-time trends of behavior, reducing dynamic models' predictive capacities [7, 27].

**Subjective Sources**: Psychobehavioral models heavily depended on subjective reports and consequently reporting biases [2, 3].

### Limitations of the Review Processes Used

Heterogeneity of Study Designs: Various research designs (i.e., supervised learning, unsupervised clustering, and neural networks) created it challenging to directly compare studies [9, 14].

Reporting Bias Sensitivity: Positive results were more likely to be reported, while negative or neutral findings could have been underreported [10, 19].

Restricted Meta-Analysis: As a result of heterogeneity of outcome and statistical approaches used, a successful meta-analysis wasn't possible to perform [4, 11].

### d) Implications of the Results for Practice, Policy, and Future Research

There are numerous implications of the results for practice, policy, and future research:

**Practice**

Institutional use can be made of GCN-based models for individualized performance monitoring and early intervention [12].

Time-based models can make it possible for teachers to develop adaptive learning systems that are consistent with student learning styles [1].

Blending behavioral data mining with learning management systems can enhance monitoring of student engagement [8].

**Policy**

Implementation of data privacy regulations for the collection of behavioral data of students is required to avoid misusage [9, 23].

There is a need for education policy to encourage explainable AI model usage to introduce transparency and fairness in predicting performance [11].

**Future Research**

Future research will involve multi-modal behavioral data integrating academic, psychological, and social activities [6, 14].

Real-time analysis of behavioral patterns using dynamic model-based analysis requires more research [27].

Offline learning behavior and the influence of extracurricular activities on student performance need to be investigated [19, 28].

Cross-cultural validation of the model will enhance validity in performance prediction models across different student groups [3, 15].

## V. REFERENCES

1. Deeva, G., De Smedt, J., Saint-Pierre, C., Weber, R., & De Weerdt, J. (2022). Predicting student performance using sequence classification with time-based windows. Expert Systems With Applications, 209, 118182. https://doi.org/10.1016/j.eswa.2022.118182

2. Qu, D., & Guan, Q. (2022). Predictive Model and Analysis of Psychological Depression Based on College Students' Behavior Data Mining. Wireless Communications and Mobile Computing, 2022, 1–10. https://doi.org/10.1155/2022/5352283

3. Li, H. (2023). The Influence of Online Learning Behavior on Learning Performance. Applied Science and Innovative Research, 7(1), p69. https://doi.org/10.22158/asir.v7n1p69

4. Data Analytics on Online Student Engagement Data for Academic Performance Modeling. (2022). IEEE Access, 10, 103176–103186. https://doi.org/10.1109/access.2022.3208953

5. Arizmendi, C. J., Bernacki, M. L., Rakovic, M., Plumley, R. D., Urban, C. J., Panter, A. T., Greene, J. A., & Gates, K. M. (2022). Predicting student outcomes using digital logs of learning behaviors: Review, current standards, and suggestions for future work. Behavior Research Methods. https://doi.org/10.3758/s13428-022-01939-9

6. Jiao, P., Ouyang, F., Zhang, Q., & Alavi, A. H. (2022). Artificial intelligence-enabled prediction model of student academic performance in online engineering education. Artificial Intelligence Review, 55(8), 6321–6344. https://doi.org/10.1007/s10462-022-10155-y

7. Su, Y. S., Lin, Y.-D., & Liu, T.-Q. (2022). Applying machine learning technologies to explore students' learning features and performance prediction. Frontiers in Neuroscience, 16. https://doi.org/10.3389/fnins.2022.1018005

8. Guo, T., & Xia, F. (2023). Student Behavior Pattern Mining and Analysis: Towards Smart Campuses. Web Search and Data Mining. https://doi.org/10.1145/3539597.3575780

9. Akhuseyinoglu, K., & Brusilovsky, P. (2022). Exploring Behavioral Patterns for Data-Driven Modeling of Learners' Individual Differences. Frontiers in Artificial Intelligence, 5. https://doi.org/10.3389/frai.2022.807320

10. Hooshyar, D., & Yang, Y. (2021). Predicting Course Grade through Comprehensive Modelling of Students' Learning Behavioral Pattern. Complexity, 2021, 1–12. https://doi.org/10.1155/2021/7463631

11. Ramaswami, G. S., Sušnjak, T., & Mathrani, A. (2022). Supporting Students' Academic Performance Using Explainable Machine Learning with Automated Prescriptive Analytics. Big Data and Cognitive Computing, 6(4), 105. https://doi.org/10.3390/bdcc6040105

12. Mubarak, A. A., Hezam, I. M., & Hao, F. (2022). Modeling students' performance using graph convolutional networks. Complex & Intelligent Systems, 8(3), 2183–2201. https://doi.org/10.1007/s40747-022-00647-3

13.Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1). https://doi.org/10.1186/s40561-022-00192-z

14. Gonzalez-Nucamendi, A., Noguez, J., Neri, L., Robledo-Rella, V., García-Castelán, R. M. G., & Escobar-Castillejos, D. (2022). Learning Analytics to Determine Profile Dimensions of Students Associated with their Academic Performance. Applied Sciences, 12(20), 10560. https://doi.org/10.3390/app122010560

15. Salazar-Fierro, F. A., García-Santillán, J., Posso-Yepez, M., Landeta-López, P. A., & García-Santillán, I. (2023). Analysis of Student Performance Applying Data Mining Techniques in a Virtual Learning Environment. International Journal of Emerging Technologies in Learning (Ijet, 18(11), 175–195. https://doi.org/10.3991/ijet.v18i11.37309

16. Hande, K. N. (2022). A survey on Prediction and Analysis of Students Academic Performance Using Machine Learning. International Journal For Science Technology And Engineering, 10(6), 780–792. https://doi.org/10.22214/ijraset.2022.43192

17.Priyambada, S. A., Usagawa, T., & Er, M. (2023). Two-layer ensemble prediction of students' performance using learning behavior and domain knowledge. Computers & Education: Artificial Intelligence, 5, 100149. https://doi.org/10.1016/j.caeai.2023.100149

18. Xu, F., & Qu, S. (2021). Data Mining of Students' Consumption Behaviour Pattern Based on Self-Attention Graph Neural Network. Applied Sciences, 11(22), 10784. https://doi.org/10.3390/APP112210784

19. Chen, L., Wang, L., & Zhou, Y. (2022). Research on Data Mining Combination Model Analysis and Performance Prediction Based on Students' Behavior Characteristics. Mathematical Problems in Engineering, 2022, 1–10. https://doi.org/10.1155/2022/7403037

20. Jenitha, T., Santhi, S., & Privthy Jeba, J. M. (2021). Prediction of Students' Performance based on Academic, Behaviour, Extra and Co-Curricular Activities. NeuroQuantology Journal, 18, 262–279. https://doi.org/10.14704/WEB/V18SI01/WEB18058

21. Bao, Y., Lu, F., Yanxiao, W., Zeng, Q., & Liu, C. (2020). Student Performance Prediction Based on Behavior Process Similarity. Chinese Journal of Electronics, 29(6), 1110–1118. https://doi.org/10.1049/CJE.2020.02.012

22. Na, W. (2020). A Data Mining Method for Students' Behavior Understanding. International Journal of Emerging Technologies in Learning (Ijet), 15(06), 18–32. https://doi.org/10.3991/IJET.V15I06.13175

23. Development of a Neural Network for modeling the behavior of the Educational Process. (2023). Iskusstvennye Obŝestva, 18(1), 0. https://doi.org/10.18254/s207751800024453-7

24. Mirzaei, M., & Sahebi, S. (2019). Modeling Students' Behavior Using Sequential Patterns to Predict Their Performance (pp. 350–353). Springer, Cham. https://doi.org/10.1007/978-3-030-23207-8_64

25. Sharma, K., Papamitsiou, Z., Olsen, J. K., & Giannakos, M. N. (2020). Predicting learners' effortful behaviour in adaptive assessment using multimodal data. Learning Analytics and Knowledge, 480–489. https://doi.org/10.1145/3375462.3375498

26. Yao, H., Lian, D., Cao, Y., Wu, Y., & Zhou, T. (2019). Predicting Academic Performance for College Students: A Campus Behavior Perspective. arXiv: Computers and Society. http://export.arxiv.org/pdf/1903.06726

27. An Unsupervised Ensemble Clustering Approach for the Analysis of Student Behavioral Patterns. (2022). International Journal for Innovative Engineering and Management Research, 481–490. https://doi.org/10.48047/ijiemr/v11/i06/29

28. Zhang, X., Sun, G., Pan, Y., Sun, H., He, Y., & Tan, J. (2018). Students performance modeling based on behavior pattern. Journal of Ambient Intelligence and Humanized Computing, 9(5), 1659–1670. https://doi.org/10.1007/S12652-018-0864-6

29. Almutairi, S., Shaiba, H., & Bezbradica, M. (2019). Predicting Students' Academic Performance and Main Behavioral Features Using Data Mining Techniques (pp. 245–259). Springer, Cham. https://doi.org/10.1007/978-3-030-36365-9_21

30. Kassarnig, V., Mones, E., Bjerre-Nielsen, A., Sapiezynski, P., Sapiezynski, P., Lassen, D. D., Lehmann, S., & Lehmann, S. (2018). Academic performance and behavioral patterns. EPJ Data Science, 7(1), 1–16. https://doi.org/10.1140/EPJDS/S13688-018-0138-8

31. Lorenzen, S. S., Hjuler, N., & Alstrup, S. (2019). Tracking Behavioral Patterns among Students in an Online Educational System. arXiv: Computers and Society. https://arxiv.org/abs/1908.08937