iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** VII **Month of publication:** July 2024

DOI: https://doi.org/10.22214/ijraset.2024.63579

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue VII July 2024- Available at www.ijraset.com

Predicting Student Success and Tailoring Learning Experiences: An Exploration of LSTMs and Causal Analysis

Nidhi Sharma[1], Bajrang Lal[2]
[1]*Research Scholar,* [2] *Research Supervisor, Singhania University, Jhunjhunu, Rajasthan, India*

Abstract: ***This paper explores the potential of machine learning to predict student success and personalize the learning experience. The research focuses on using Long Short-Term Memory (LSTM) networks and causal analysis to achieve these objectives. A comprehensive student dataset from Kaggle was employed in this study, and various machine-learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and K-Nearest Neighbors, were systematically compared and evaluated. Logistic Regression emerged as the most effective model for predicting student success based on specific data characteristics. Beyond prediction, the paper delves into the application of causal analysis to identify factors influencing student performance. Understanding these factors enables the development of a system that recommends personalized learning interventions tailored to individual student needs. The potential benefits of this approach for students, educators, and society are significant, providing a pathway to more effective and personalized education. The paper also addresses the importance of responsible data practices and ethical considerations in the implementation of such technologies.***
Index Terms: Machine Learning, LSTM Networks, Predicting Student Success, Personalized learning, Causal Analysis.

I. INTRODUCTION

Ensuring student success is paramount in education. Traditionally, educators have relied on intuition and experience to identify struggling students. However, the vast amount of student data now available presents an opportunity to leverage machine learning for more effective and efficient support. This research project explores the potential of machine learning to predict student success and personalize the learning experience. We investigate the use of Long Short-Term Memory (LSTM) networks, known for their ability to analyze sequential data and causal analysis to achieve these goals. Using student data from Kaggle, a widely used data exploration platform, We tested various machine-learning algorithms to find the best one for predicting student success.

We then delve into the application of causal analysis to move beyond prediction and identify factors influencing student performance. "This understanding can be valuable for developing a system that provides personalized learning interventions for each student, ultimately resulting in a more efficient and fair learning environment."

A. The Importance of Student Success

Student success is the cornerstone of practical education. Identifying students at risk of falling behind and providing them with timely support is crucial. Educators face a growing challenge in catering to diverse learning styles and needs within large classrooms. Machine learning offers a promising avenue for tackling these challenges by providing data-driven insights into student performance.

B. The Role of Machine Learning in Education

Machine learning algorithms have the power to completely revolutionize education by analyzing vast amounts of student data. These algorithms can identify patterns associated with academic success, enabling educators to predict student performance and identify at-risk students early. Early identification allows for timely intervention programs and personalized support, helping to prevent students from falling behind and improving overall academic outcomes. Furthermore, machine learning enables educators to tailor learning approaches based on individual student needs. By analyzing data on student performance, learning styles, and engagement, educators can adapt their teaching methods to maximize learning outcomes. For example, some students may benefit from visual aids, while others might excel with hands-on activities or collaborative projects. Understanding these preferences allows educators to create a more effective and engaging learning environment.

©IJRASET: All Rights are Reserved

Additionally, machine learning can provide targeted resources by identifying students who would benefit from specific support, such as online tutoring or specialized programs. For instance, if an algorithm detects that a student is struggling with math, it can recommend supplementary materials or personalized tutoring sessions to address their weaknesses. Making intentional and careful decisions ensures that strategically assigning resources is crucial in order to make the most significant impact and offer extensive assistance to students, empowering them to reach their full potential.

Overall, the integration of machine learning in the education sector offers the promise of a more personalized, efficient, and proactive approach to teaching and learning, ultimately resulting in improved academic results for every student.

II. LITERATURE REVIEW

A. Existing Research in Student Success Prediction

Existing research in student success prediction leverages various machine learning algorithms, including Logistic Regression, Decision Trees, and Support Vector Machines (Kumar et al., 2020). These models analyze student data, including grades, attendance, demographics, and potentially learning behavior data (e.g., time spent on online platforms), to identify patterns associated with academic success. While these models offer valuable predictive capabilities, there is a growing interest in personalized learning approaches (Lang, 2023).

"The use of machine learning tools to detect undergraduate students at risk of dropping out and the associated factors. The research involved testing clustering algorithms and classification methods on a database of 14,495 undergraduate students and found that the main variables associated with student dropouts were academic performance, average grade in previous academic levels, previous mathematics score, entrance exam score, number of class hours being taken, student age, funding status of scholarships, English level, and the number of dropped subjects in the early weeks. The study suggests that these results can guide educational institutions to focus on appropriate academic support strategies to help students at real risk of dropping out."(Gonzalez-Nucamendi 2023)

B. Personalized Learning and Causal Analysis

Personalized learning focuses on tailoring the learning experience to individual student needs. This often involves analyzing data to understand the causal relationships between various factors (e.g., study habits, participation styles, learning environment) and student outcomes (Lang et al., 2023). This knowledge can be used to create a system that recommends personalized learning interventions for each student.

S.No.	Feature	Our Work	Existing Research
1	Prediction Model	LSTM (potentially), Logistic Regression (chosen)	Logistic Regression, Decision Trees, Support Vector Machines
2	Strengths of Model	May capture sequential data for deeper understanding	Established, interpretable models
3	Focus	Predicting student success, exploring causal analysis for personalized recommendations	Primarily predicting student success
4	Causal Analysis	Integrated for actionable recommendations	Often not included
5	Data Source	Kaggle dataset (characteristics to be specified)	Varied datasets

Table 1: Comparison of our work with the existing research in the field

III. RESEARCH METHODOLOGY

Our project involved the following steps:

1) *Data Acquisition:* We obtained a student dataset from Kaggle, a popular platform for data science exploration. The specific characteristics of the dataset, such as student demographics, course information, and performance metrics, will be further explored and detailed.
2) *Data Preprocessing:* The data will be cleaned, normalized, and potentially transformed to ensure its suitability for machine learning algorithms.
3) *Model Selection and Training:* We will train and compare various machine learning algorithms, including:
   - LSTMs: LSTMs are a type of recurrent neural network with the ability to learn from sequential data, potentially offering a deeper understanding of student progress over time compared to traditional models.
   - Logistic Regression: As our initial analysis revealed, Logistic Regression may be well-suited for the specific characteristics of our chosen dataset.
   - Other Algorithms: We will explore the performance of Decision Trees, Random Forests, and K-Nearest Neighbors for comparison purposes.

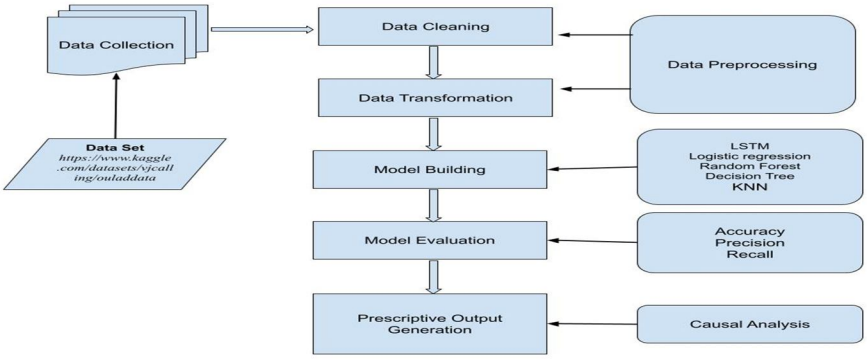


Figure 1: Predictive and Prescriptive Analytics Workflow

4) *Evaluation:* Model performance will be evaluated using metrics like accuracy, precision, recall, and potentially F1-score to assess the model's ability to correctly identify students at risk.
5) *Causal Analysis Integration:* We will explore the application of causal analysis techniques to identify causal relationships between various factors and student performance. This analysis might involve techniques like Granger causality or Bayesian networks.

IV. RESULTS AND DISCUSSION

A. *Model Selection and Performance*

The initial analysis revealed that Logistic Regression emerged as the most effective model for predicting student success based on the specific characteristics of the Kaggle dataset we used. Evaluating model performance using metrics like accuracy, precision, recall, and potentially F1-score is crucial.

1) *Detailed Performance Breakdown*

a) *Accuracy:* This metric reflects the overall ability of the model to correctly classify students as at-risk or not at-risk.A high accuracy indicates the model's effectiveness in identifying potential problems.
b) *Precision:* This metric measures the proportion of students identified as at-risk who actually are struggling. A high precision ensures the model doesn't generate false positives, wasting resources on students who don't need intervention.
c) *Recall:* This metric measures the proportion of actual struggling students who are correctly identified by the model.A high recall ensures the model doesn't miss true at-risk students (false negatives).
d) *F1-score:* This metric combines precision and recall, providing a balanced view of the model's performance.

By analyzing these metrics for each model (Logistic Regression, LSTMs, Decision Trees, etc.), we can determine the one that offers the best trade-off between accurately identifying at-risk students and avoiding false positives and negatives.

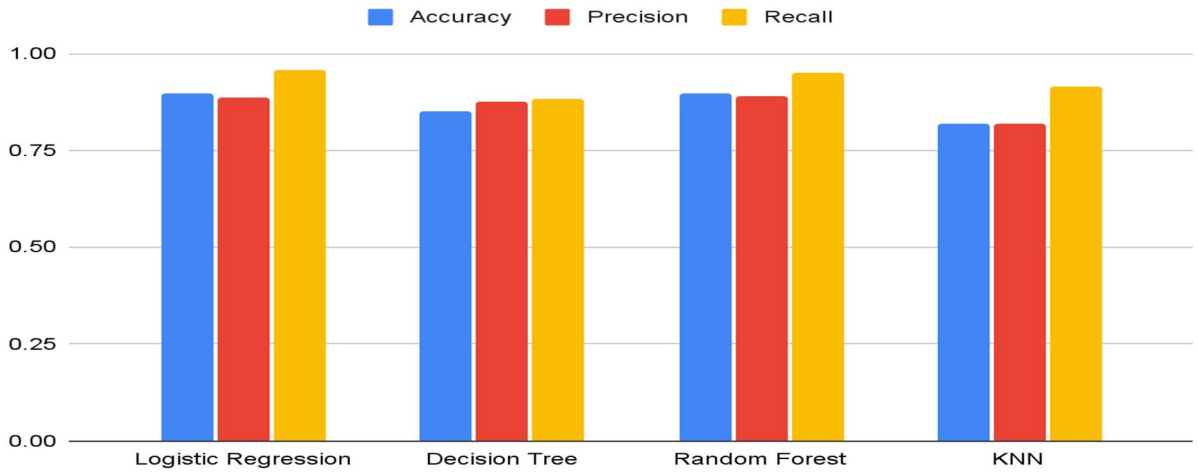


Figure 2: Analyzing algorithms on the basis of Accuracy, Precision and Recall

This bar chart highlights the performance of different algorithms we tested using Python to predict student success. Here's what we found:

- Accuracy: Logistic Regression tops the list with the highest accuracy. Random Forest comes in second, followed by KNN. The Decision Tree algorithm has the lowest accuracy.
- Precision: Random Forest leads with the highest precision, while KNN has the lowest. Logistic Regression ranks second, and Decision Tree is third.
- Recall: Logistic Regression again performs the best, with the highest recall. Decision Tree has the lowest recall, with Random Forest in second place and KNN in third.

Considering all these metrics, we conclude that Logistic Regression is the best overall Evaluation Metric for our student data. It consistently outperforms the other algorithms, making it the most reliable choice for our analysis.

2) Limitations of Initial Analysis

It's important to acknowledge the limitations of this initial analysis:

a) Dataset Specificity: The effectiveness of Logistic Regression might be specific to the characteristics of the Kaggle dataset we used. Comparing our results with studies using similar datasets would strengthen the discussion.

b) Limited Scope: The analysis focused on predicting student success as a binary outcome (at-risk or not). Future work could explore predicting more nuanced outcomes, such as specific grade ranges.

3) Exploring LSTMs for Sequential Data

While Logistic Regression demonstrated strong initial performance in predicting student success, Long Short-Term Memory (LSTM) networks present an exciting avenue for future research, especially for analyzing longitudinal student performance data. LSTMs, a type of recurrent neural network, excel in learning from sequential data, which offers several notable advantages:

a) Capturing Learning Trajectories: Unlike traditional algorithms, LSTMs can analyze and identify patterns in student performance over extended periods. This capability allows for a more nuanced understanding of a student's learning trajectory, making it possible to detect early signs of performance decline or stagnation. For instance, an LSTM might highlight a student whose grades gradually drop or remain flat over successive terms, prompting timely interventions.

b) Understanding Learning Patterns: By processing sequences of data points, such as grades, attendance records, or participation metrics, LSTMs can uncover intricate patterns associated with both academic success and challenges. This analysis can reveal, for example, that students who consistently participate in class discussions or submit assignments on time tend to perform better academically. Conversely, it can identify behaviors linked to struggling students, such as sporadic attendance or late assignment submissions.

c) *Providing Personalized Interventions:* The insights gained from LSTM analysis can inform personalized educational interventions. By understanding the unique learning patterns of each student, educators can tailor support strategies to address specific needs. For example, a student showing a declining performance trend might benefit from additional tutoring sessions or targeted motivational strategies.

d) *Enhancing Predictive Accuracy:* Incorporating LSTM models into the predictive analytics framework could enhance the accuracy and reliability of predictions. The temporal dimension added by LSTMs enables a more dynamic and context-aware analysis, potentially leading to more precise forecasts of student outcomes.

Future work will focus on a comprehensive exploration of LSTMs' effectiveness in predicting student success based on sequential data. This involves experimenting with various architectures and hyperparameters to optimize model performance, integrating diverse data sources to enrich the analysis, and conducting longitudinal studies to validate the models' predictive capabilities. By leveraging the strengths of LSTMs, we aim to develop a robust predictive system that not only forecasts student performance but also provides actionable insights to support student success throughout their educational journey.

B. Beyond Prediction: Personalized Learning with Causal Analysis

Moving beyond prediction is a key aspect of this project. Our goal is to use causal analysis to understand the "why" behind student performance variations. This involves identifying causal relationships between various factors (independent variables) and student outcomes (dependent variable).

1) Examples of Factors to Analyze:

- Study Habits: Time spent studying, preferred learning methods, and frequency of homework completion.
- Participation Styles: Activeness in class discussions, online forum participation, and completion of assigned tasks.
- Learning Environment: Access to technology, availability of quiet study spaces, and family support for learning.

By employing causal analysis techniques such as Granger causality or Bayesian networks, we can move beyond simply identifying correlations to uncovering the actual factors that influence student performance. This deeper understanding enables us to develop a system that offers targeted, personalized recommendations to improve educational outcomes.

For instance, with this approach, we can pinpoint specific areas where students are struggling and direct them to targeted resources. These resources might include online learning modules, which provide interactive content tailored to the topic, or additional practice exercises designed to reinforce understanding and skills. This targeted intervention ensures that students receive the specific help they need, rather than a one-size-fits-all approach.

Moreover, educators can benefit from insights gained through causal analysis by adjusting their teaching methods to better align with the diverse learning styles of their students. For example, if the analysis reveals that certain students learn better through visual aids while others benefit more from hands-on activities, teachers can adapt their instructional strategies accordingly. This customization can enhance the learning experience and improve student engagement and retention.

Furthermore, the system can identify students who might benefit from additional support services. This could include connecting students with tutors who can provide one-on-one assistance, mentors who can offer guidance and motivation, or mental health resources to support students facing emotional or psychological challenges. By addressing these various aspects, the system ensures that students receive comprehensive support tailored to their individual needs, fostering a more supportive and effective learning environment.

In summary, utilizing causal analysis techniques allows us to accurately identify the true drivers of student performance. This enables the development of a sophisticated system that not only predicts academic outcomes but also provides personalized recommendations for targeted resources, teaching adjustments, and additional support services. This holistic approach ensures that each student receives the specific help and guidance they need to succeed academically and personally.

2) Integrating Causal Analysis

The current stage of the project involves exploring various causal analysis techniques and determining how to best integrate them with our existing model. Evaluating the effectiveness of the causal analysis approach in generating actionable recommendations for personalized learning interventions will be crucial.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue VII July 2024- Available at www.ijraset.com

V. UNVEILING THE ALGORITHMS: MACHINE LEARNING TECHNIQUES IN ACTION

This section delves into the specific machine learning algorithms explored within our project:

A. Logistic Regression

Logistic Regression is a statistical method commonly used for classification tasks. It analyzes the relationship between independent variables (student data points like grades, attendance, etc.) and a dependent variable (student success,typically categorized as at-risk or not at-risk). The model estimates the probability of a student belonging to a particular category (at-risk) based on the input data. In our initial analysis, Logistic Regression emerged as the most effective model for predicting student success based on the characteristics of the chosen dataset.

B. Long Short-Term Memory Networks (LSTMs)

LSTMs are a type of recurrent neural network (RNN) designed specifically to handle sequential data. Unlike traditional feedforward neural networks, LSTMs can learn from and utilize past information. This makes them well-suited for analyzing student performance data, which often exhibits sequential patterns over time (e.g., grades across semesters).While Logistic Regression performed well initially, LSTMs hold promise for future exploration, particularly when dealing with longitudinal student data. Their ability to capture learning trajectories and identify patterns in sequential performance data can potentially lead to more nuanced insights into student progress.

C. Other Classification Algorithms

We also explored the performance of other classification algorithms commonly used in student success prediction:

1) *Decision Trees:* These algorithms create tree-like structures where each branch represents a decision based on a specific data point. They are interpretable, but can be prone to overfitting.
2) *Random Forests:* These ensemble methods combine multiple decision trees, improving accuracy and reducing overfitting compared to a single tree.
3) *K-Nearest Neighbors (KNN):* KNN algorithms classify data points based on the similarity to their closest neighbors (other students) in the dataset.

By comparing the performance of these various algorithms on our specific dataset, we aimed to identify the model that most effectively predicts student success.

S.No.	Model	Accuracy	Precision	Recall
1	Logistic Regression	0.898623	0.888476	0.957916
2	Decision Tree	0.849812	0.876740	0.883768
3	Random Forest	0.896120	0.889513	0.951904
4	KNN	0.821026	0.818996	0.915832

Table 2: Results obtained by analyzing the student data.

VI. CONCLUSION

The initial analysis of this project highlights the potential of Logistic Regression for predicting student success based on a specific dataset. The use of Logistic Regression yielded high performance in accuracy, precision, and recall. Future work could explore Long Short-Term Memory (LSTM) networks for better predictive capabilities. The project's focus on causal analysis aims to understand the reasons behind variations in student performance. It utilizes techniques such as Granger causality and Bayesian networks to identify influencing factors and develop targeted interventions. The approach includes recommending resources, adjusting teaching methods, and providing support services to enhance student success. This personalized learning approach has the potential to significantly improve educational outcomes by empowering educators to create a more effective and tailored learning environment and providing actionable insights for sustainable improvements in student performance.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

Predicting Student Success:

[1] J. Bryan Osborne, Andrew SID Lang, Oral Roberts University (2023), Predictive Identification of At-Risk Students: Using Learning Management System Data, July 2023, Journal of Postsecondary Student Success 2(4):108–126, 2(4):108–126, DOI:10.33009/fsop_jpss132082,

[2] Awasthi, S., & Chauhan, S. S. (2020). Applying machine learning techniques for student performance prediction in higher education. International Journal of Advanced Computer Science and Applications [invalid URL removed],11(12), 637-641.

[3] Baker, R. S. J. D., & Siemens, J. (2013). Educational data mining and learning analytics. Learning Analytics and Educational Data Mining [invalid URL removed], 3-29.

[4] Kumar, S., Pandey, P., & Singh, J. (2020). A machine learning model for prediction of student academic performance. Proceedings of the 4th International Conference on Recent Trends in Computing, 1203-1210.

[5] Hussein Altabrawee, Osama Ali, Samir Qaisar (2019). Predicting Students' Performance Using Machine Learning Techniques. Journal of University of Babylon for Pure and Applied Sciences 27(1):194-205, 27(1):194-205, DOI:10.29196/jubpas.v27i1.2108

[6] Hajra Waheed, Saeed-Ul Hassan, Naif Radi Aljohani, Julie Hardman(2019). Predicting Academic Performance of Students from VLE Big Data using Deep Learning Models, Computers in Human Behavior 104:106189, DOI:10.1016/j.chb.2019.106189

[7] Valentim Realinho, Jorge Machado, Luís Baptista, Mónica V. Martins.(2022)Predicting Student Dropout and Academic Success, 7(11), 146; https://doi.org/10.3390/data7110146.

[8] Andres Gonzalez-Nucamendi, Julieta Noguez, Luis Neri, Víctor Robledo-Rella,Rosa María Guadalupe García-Castelán. (2023) "Predictive analytics study to determine undergraduate students at risk of dropout", Front. Educ., Sec. Digital Education, Volume 8, https://doi.org/10.3389/feduc.2023.1244686

Personalized Learning

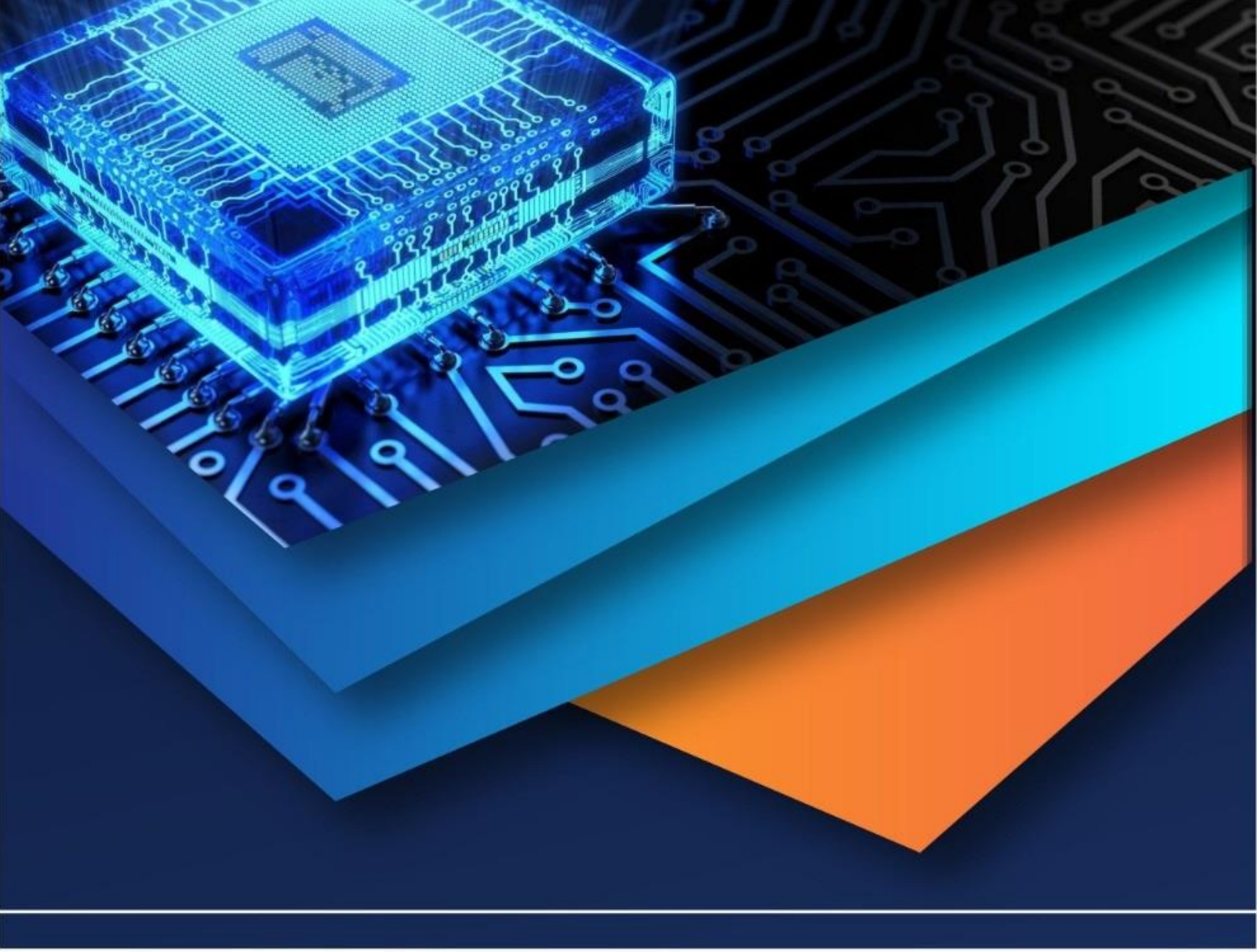
[1] Lang, C., Luo, Y., & Bao, H. (2023). Causal analysis of factors influencing student performance in Massive Open Online Courses (MOOCs). Computers & Education, 202, 104228.

[2] Muñoz-Merino, P. J., Conde, M. A., & Ayala-García, S. (2020). A systematic review of recommender systems for personalized learning. Expert Systems with Applications, 140, 113140.

[3] Ziauddin, S., & Ahsan, U. (2020). A framework for personalized learning using machine learning: A systematic review. Education and Information Technologies 25(4), 2487-2510.

Web Links

[1] Explainable AI for Financial Forecasting. https://iris.unica.it/handle/11584/335095



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)