

# AI-Based Student Dropout Prediction System

Tanisha Jitendra Kadam <sup>1</sup>, Aarti Sanjay Gawai <sup>2</sup>

<sup>1</sup> Student, Department of Information Technology and Computer Science, D.G. Ruparel College

<sup>2</sup> Assistant Professor, Department of Information Technology and Computer Science, D.G. Ruparel College, Mumbai, Maharashtra, India

<sup>1</sup> kadamtanisha589@gmail.com

<sup>2</sup> aartigawai1611@gmail.com

\*\*\*

**Abstract** - Student dropout is a concern for colleges and universities because some students leave their studies before completing their course. Finding such students early may give institutions more time to understand their difficulties and offer suitable help. This paper studies the use of Artificial Intelligence (AI) and Machine Learning (ML). The study is based on previously published research, reports, and available datasets. It looks at factors that may be connected with dropout, including marks, attendance, earlier academic records, student background, and participation in learning activities. Common ML methods such as Decision Tree, Random Forest, Logistic Regression, Support Vector Machine, and Artificial Neural Networks are considered in the review. Measures including accuracy, precision, recall, F1-score, and AUC are used in earlier studies to judge prediction performance. The paper also looks at Explainable AI methods such as SHAP and LIME, which can help show the factors behind a prediction. The findings suggest that AI and ML can be useful for finding possible dropout risk before the student leaves. However, incomplete data, unequal class distribution, fairness, and differences between institutions can affect the results. Therefore, prediction should be used to guide student support rather than to label a student as certain to drop out.

**Key Words:** Student dropout, Artificial Intelligence, Machine Learning, Dropout Prediction, Explainable AI, Student Retention

## 1. INTRODUCTION

Nowadays, worldwide, one of the major challenges faced by educational institutions is Student Dropout. When a student enrolls in an educational program but they leave that program before completing the required course or degree, student dropout occurs. Student Dropout can negatively affect the student's academic and career opportunities, and it can also create challenges for educational institutions. Identifying students who do not continue their education at an early stage is important so that the student institute can provide appropriate support and guidance.

Traditional methods of identifying students at risk of dropping out often depend on academic performance, attendance records, manual analysis, and teacher observations. These methods may not always be effective

because large amounts of student information must be examined, and warning signs may be identified only after the student has already started experiencing difficulties. There is a need for more efficient and data-driven approaches that can identify potential dropout risks at an early stage.

Artificial Intelligence (AI) and Machine Learning (ML) provide techniques that can analyze large amounts of student-related data and identify patterns associated with student dropout. Factors such as attendance, examination results, academic performance, previous academic history, social and financial conditions, and other relevant characteristics can be analyzed to estimate the likelihood of a student dropping out. Machine Learning models can learn from historical student data and use these patterns to classify students as being at low, medium, or high risk of dropout.

An AI-based Student Dropout Prediction System can act as an early-warning mechanism for educational institutions. The purpose of such a system is not only to determine with certainty that a student will drop out, but also to identify students who may require additional attention or support. Institutions can then take appropriate preventive measures, such as academic counselling, mentoring, attendance monitoring, financial assistance, or other forms of student support.

This research focuses on studying the application of Artificial Intelligence and Machine Learning techniques for predicting student dropout. It aims to examine the factors that contribute to student dropout, investigate existing prediction approaches, and understand how AI-based prediction can support early intervention. The study can help highlight the potential benefits as well as the limitations and challenges of using AI for student dropout prediction, including data quality, fairness, accuracy, and privacy.

## 2. REVIEW OF LITERATURE

Student Dropout Prediction has become an important research area in educational data mining because early identification of at-risk students can help institutions provide timely support. Existing studies have investigated dropout using academic, demographic, socioeconomic, behavioral, and institutional factors.

Rodrigues et al. [1] reviewed artificial intelligence algorithms used for predicting student dropout and highlighted the growing use of AI-based approaches in student retention.

Silva and Roman [2] systematically reviewed dropout prediction methods in higher education and demonstrated the importance of data-driven approaches for identifying students at risk.

Giron et al. [3] reviewed machine-learning and deep-learning techniques for dropout prediction and reported the use of models such as Decision Tree, Logistic Regression, Random Forest, and deep-learning methods. Agrusti et al. [4] demonstrated the application of deep learning for university dropout prediction, while Sultan et al. [5] and Anjali and Bai [6] investigated deep-learning approaches for predicting dropout in online learning environments using learner activity data.

Pathiranagama et al. [7] examined interpretable machine learning for student dropout prediction and highlighted the usefulness of techniques such as SHAP and LIME for understanding model predictions.

Perez et al. [8] and Goren et al. [9] further demonstrated the potential of machine learning for early identification of students at risk of dropout.

Ennibras et al. [10] emphasized the potential of AI-based approaches for supporting dropout prevention and early intervention.

Lopez-Munoz et al. [11] reviewed machine-learning methods and dropout risk factors and highlighted the continued importance of structured educational data and suitable prediction algorithms.

### Research Gap

Although existing studies demonstrate the effectiveness of artificial intelligence, machine learning, and deep learning for dropout prediction. Many studies focus mainly on prediction accuracy and use datasets from individual institutions, which can limit generalizability. There is also a need for greater explainability and consideration of multiple student-related factors.

Furthermore, prediction alone does not prevent dropout; the results should support timely and meaningful intervention. Finally, there remains a gap between predicting dropout risk and using prediction results for effective early intervention. Existing research provides evidence that AI can identify potentially at-risk students, but further work is required to ensure that these predictions can support meaningful and timely educational interventions.

Therefore, the proposed research on an AI-Based Student Dropout Prediction System aims to address these limitations by investigating suitable artificial-intelligence and machine-learning approaches for identifying students at risk of dropout, considering relevant student characteristics, and providing interpretable prediction results. The study focuses on the potential of AI-based prediction as a decision-support mechanism that can assist educational institutions in identifying at-risk

students at an early stage and supporting student-retention efforts.

| Research Gap          | Existing Problem                                                                             | Possible Research Direction                                                                      |
|-----------------------|----------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Student Data          | Student's Available data may be incomplete.                                                  | Use better-quality and varied student information.                                               |
| Model Transferability | A model may give different results with students from another institution.                   | Test the model with different student datasets                                                   |
| Model Explanation     | Some predictions do not show what caused the risk.                                           | To identify the factors which are behind each prediction for that, we can use the XAI technique. |
| Prediction Fairness   | Based on the previous study for different student groups, the model may perform differently. | One model can perform similarly for all student groups.                                          |
| Student Support       | Identifying a high-risk student does not itself solve the dropout problem.                   | Study how predictions can guide early and suitable support.                                      |

**Table 1. Research Gap Comparison**

### 3. PROBLEM IDENTIFICATION

Student Dropout is a major concern for higher-education institutions because leaving a course before completion can affect a student's academic progress and future opportunities. In institutions mainly dropout reasons are academic performance, attendance, engagement, financial conditions, or the learning environment. Identifying students who may discontinue their education at an early stage is therefore important for providing suitable support. In many educational institutions, the identification of student is at risk of dropout mainly depends upon academic records, attendance, teacher observations, or manual monitoring. These approaches can become difficult when the number of students is large, and information is stored across different records. A student may also show a combination of risk indicators that are difficult to recognize through manual observation alone. Machine-learning techniques provide an opportunity to analyze student data and discover patterns associated with dropout. However, previous prediction

techniques have several limitations. Some studies concentrate mainly on academic performance and may not sufficiently consider other relevant student characteristics. In addition, models developed using data from a particular institution may not perform equally well in different educational settings. Another concern is the interpretation of prediction results. A system may classify a student as having a high probability of dropping out, but simply providing this classification may not be sufficient for educational staff. We may also need to understand which technique is useful for the prediction. Without understandable results, it can be difficult to use an automated prediction system confidently for student support. The class imbalance problem is another challenge in dropout prediction. In many systems, the pursuing students are examined more than the dropout students. Sequentially, a system failed to identify correctly to that student who is actually at risk. There is also a practical gap between predicting dropout and preventing it. Identifying a high-risk student is only the first step. The prediction should provide useful information that enables educators or institutions to consider appropriate support at an early stage. Therefore, a useful dropout prediction system should focus not only on classification performance but also on understandable results and early identification. Based on these issues, the main problem identified in this research is the difficulty of reliably identifying students who are likely to drop out before the problem becomes severe. The proposed research therefore focuses on investigating an AI-based approach that can analyze relevant student information, early identify potential dropout risk, and present the prediction in an understandable form.

#### 4. OBJECTIVES OF THE STUDY

- i. To understand the different academic and non-academic conditions that may increase the possibility of a student leaving higher education before completing the course.
- ii. To study the role of factors such as academic results, attendance, personal background, engagement, and previous educational performance in student dropout.
- iii. To review the Machine Learning approaches that have been applied in previous studies for identifying students with a higher likelihood of dropping out.
- iv. To understand the differences between commonly used prediction models and determine which approaches are more appropriate for student dropout analysis.
- v. To assess the predictive ability of selected Machine Learning models through measures including accuracy, precision, recall, F1-score, and AUC.
- vi. To study how historical student information can be analysed to recognize warning signs before a student actually drops out.
- vii. To explore Explainable AI methods that can provide understandable reasons behind a model's prediction instead of presenting only a dropout risk result.
- viii. To understand which student-related characteristics have the greatest influence on the predicted risk of dropout.
- ix. To examine how the findings from an AI-based prediction approach could assist educational institutions in planning suitable support for students who may be experiencing difficulties.
- x. To determine the major limitations and unresolved research issues in existing AI-based student dropout prediction studies and identify areas that require further investigation.

#### 5. HYPOTHESIS OF THE STUDY

The study examines whether student-related characteristics and Artificial Intelligence (AI) and Machine Learning (ML) techniques can contribute to the prediction of student dropout in higher education. Based on the objectives of the study, the following null and alternative hypotheses are proposed.

H<sub>01</sub>: Academic performance, attendance, demographic characteristics, and behavioral patterns are not significantly associated with student dropout.

H<sub>11</sub>: Academic performance, attendance, demographic characteristics, and behavioral patterns are significantly associated with student dropout.

H<sub>02</sub>: Machine Learning techniques do not provide better results in predicting student dropout than conventional prediction methods.

H<sub>12</sub>: Machine Learning techniques provide better results in predicting student dropout than conventional prediction methods.

H<sub>03</sub>: There is no significant difference in the predictive performance of different Machine Learning algorithms for student dropout prediction.

H<sub>13</sub>: There is a significant difference in the predictive performance of different Machine Learning algorithms for student dropout prediction.

H<sub>04</sub>: Explainable Artificial Intelligence does not provide significant additional understanding of the factors influencing student dropout predictions.

H<sub>14</sub>: Explainable Artificial Intelligence provides significant additional understanding of the factors influencing student dropout predictions.

H<sub>05</sub>: Early identification of students at risk of dropping out does not significantly support timely intervention by educational institutions.

H<sub>15</sub>: Early identification of students at risk of dropping out significantly supports timely intervention by educational institutions.

## 6. RESEARCH DESIGN AND METHODOLOGY

This study follows a descriptive and analytical research design based mainly on secondary data. Relevant research papers, journal articles, conference papers, reports, and publicly available datasets related to student dropout prediction are reviewed. The study focuses on identifying dropout-related factors, examining Machine Learning techniques, comparing reported findings, and understanding the role of Explainable Artificial Intelligence.

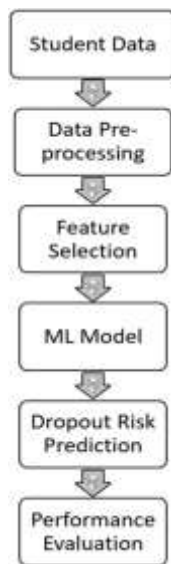


Fig.1 Research Design

The Study examines different factors associated with student dropout.

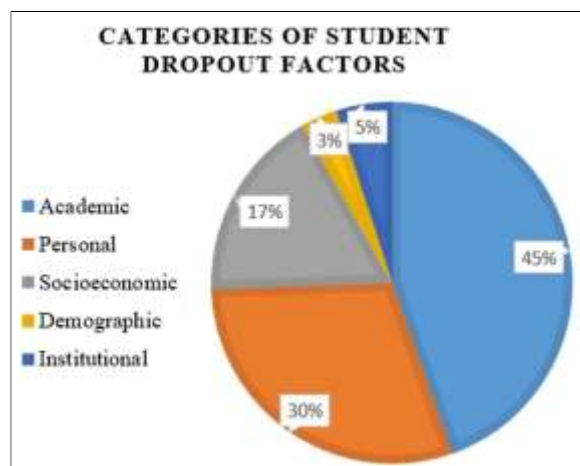


Chart 1. Categories of student dropout factors

Different Machine Learning techniques, including Decision Tree, Random Forest, Logistic Regression, SVM, and Artificial Neural Networks, are reviewed and compared based on their reported performance.

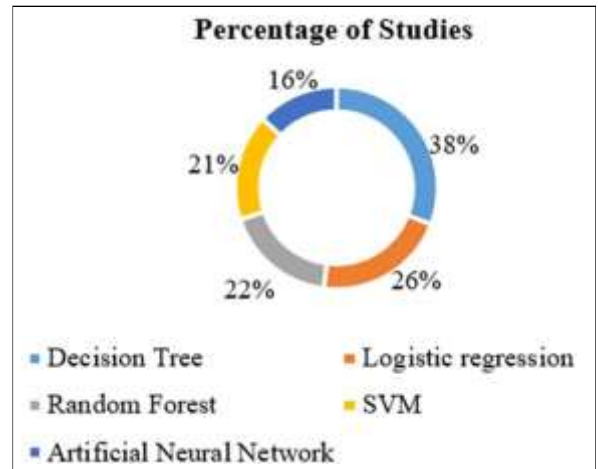


Chart 2. Use of Machine Learning Models

The performance of prediction models is commonly evaluated using measures such as Accuracy, Precision, Recall, F1-score, and AUC. These measures help determine how effectively a model identifies students who are at risk of dropping out.

| Measure   | Purpose                                               |
|-----------|-------------------------------------------------------|
| Accuracy  | Measures overall correct prediction                   |
| Precision | Measures the correctness of positive predictions      |
| Recall    | Measures the ability to identify actual at-risk cases |
| F1-score  | Combines Precision and Recall                         |

Table 2. Common Evaluation Measures

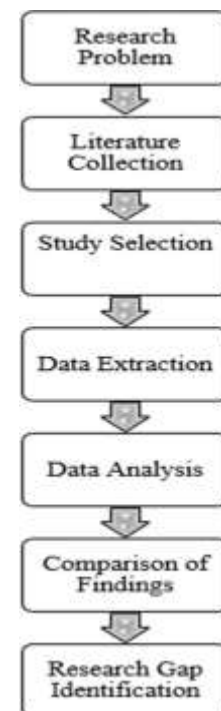


Fig 2. Machine Learning Prediction Process

Explainable Artificial Intelligence is considered important because prediction results should be understandable to educational institutions. The systematic review identified eight XAI models, with SHAP and LIME being the most widely used approaches.

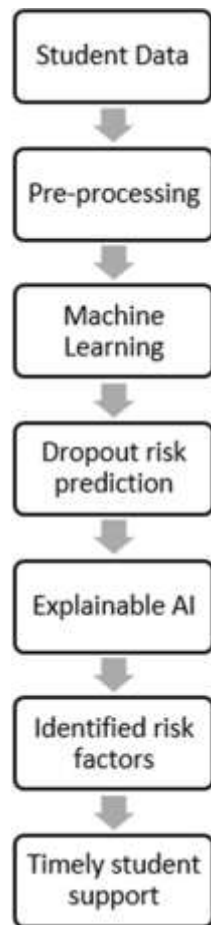


Fig 3. Explainable AI Process

Based on the reviewed literature, a conceptual framework is proposed in which student information is processed through a Machine Learning model to estimate dropout risk. Explainable AI can then be used to identify the factors contributing to the prediction, supporting timely institutional assistance.

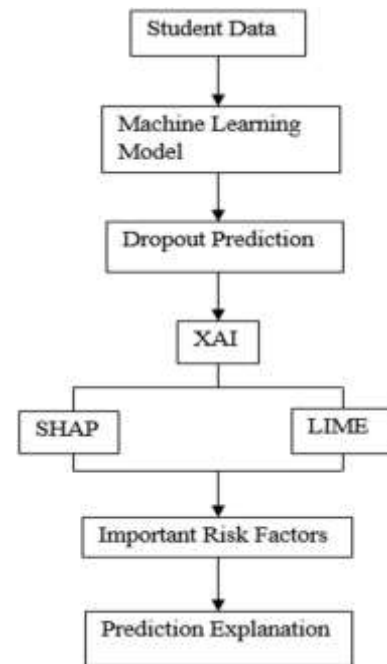


Fig 4. Proposed AI-Based Student Dropout Prediction Framework

## 7. CONCLUSIONS

Student dropout is a serious issue for higher education institutions. From the studies reviewed, it can be seen that Artificial Intelligence and also the Machine Learning can help in finding students who may be at risk of leaving their studies. Information such as marks, attendance, previous academic records, and student behavior can be useful for making these predictions.

Different ML methods, such as Decision Tree, Random Forest, Logistic Regression, SVM, and Artificial Neural Network, have been used for finding students who may be at risk of leaving their studies. Student performance can be checked using measures such as the accuracy of the student attendance, precision, recall, F1-score, and AUC. Explainable AI methods such as SHAP and LIME can also help in understanding the reasons behind a prediction.

However, issues such as missing data of student, class imbalance, fairness, and differences between institutions still need attention. Also, predicting a student at risk does not by itself prevent dropout. The prediction should help teachers and institutions provide suitable support at the right time.

Thus, AI-based student dropout prediction can be useful as an early-warning approach. Future studies should focus on improving prediction reliability, making the results easier to understand, and connecting predictions with effective student-support activities.

## ACKNOWLEDGEMENT

I would like to thank everyone who supported me in completing this research paper on "AI-Based Student Dropout Prediction System."

I am thankful to my research guide for the guidance, suggestions, and support given during this research. I also thank my teachers and college for providing the necessary academic support and resources.

I would like to acknowledge the researchers and authors of the studies referred to in this paper. Their research work helped me understand the different factors related to student dropout, Machine Learning techniques, prediction methods, and Explainable AI.

I am also thankful to my family and friends for their support and encouragement throughout the research work.

Finally, I thank everyone who helped me directly or indirectly in completing this research paper.

## REFERENCES

- [1] Rodrigues, H., Santiago, E., Wanderley, G., et al.: Artificial Intelligence Algorithms to Predict College Students' Dropout: A Systematic Mapping Study. In: International Conference on Agents and Artificial Intelligence (2024). doi:10.5220/0012348000003636
- [2] Silva, J., Roman, N.T.: Predicting Dropout in Higher Education: A Systematic Review. In: Anais do XXXII Simpósio Brasileiro de Informática na Educação (2021). doi:10.5753/sbie.2021.217437
- [3] Girón, D.A., Rosas, J.S., Marín-Rodríguez, W., et al.: Predicting Student Dropout Based on Machine Learning and Deep Learning: A Systematic Review. EAI Endorsed Transactions on Scalable Information Systems (2023). doi:10.4108/eetsis.3586
- [4] Agrusti, F., Mezzini, M., Bonavolontà, G.: Deep Learning Approach for Predicting University Dropout: A Case Study at Roma Tre University (2020). doi:10.20368/1971-8829/1135192
- [5] Sultan, M., Elsayed, H., Khan, M., Abduljabar, M.: A Deep Learning Model for MOOC Dropout Prediction Using Learner's Course-Relevant Activities. In: Global Conference on Artificial Intelligence and Internet of Things (2022). doi:10.1109/GCAIoT57150.2022.10019063
- [6] Anjali, C., Bai, V.R.: An Early Prediction of Dropouts for At-Risk Scholars in MOOCs Using Deep Learning. In: 2022 International Conference on Computing, Communication, Security and Intelligent Systems (2022). doi:10.1109/IC3SIS54991.2022.9885328
- [7] Pathiranagama, M., Nanayakkara, I., Abeyweera, R., Halloluwa, T.: Beyond the Black Box: An Interpretable Machine Learning Approach to Student Dropout Prediction. In: International Conference on Automation and Computing (2025). doi:10.1109/ICAC69156.2025.11361534
- [8] Pérez, M., Navarrete, D., Calisto, M.G.B., et al.: Unlocking Student Success: Applying Machine Learning for Predicting Student Dropout in Higher Education. In: International Symposium on Digital Forensics and Security (2025). doi:10.1109/ISDFS65363.2025.11012013
- [9] Goren, O., Cohen, L., Rubinstein, A.: Early Prediction of Student Dropout in Higher Education Using Machine Learning Models. In: Educational Data Mining (2024).
- [10] Ennibras, F., Aoula, E., Bouihi, B.: AI in Preventing Dropout in Distance Higher Education: A Systematic Literature Review. In: 2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology (2024). doi:10.1109/IRASET60544.2024.10548954
- [11] Lopez-Muñoz, G., Serrano, C.G., Sanchez-Ferreira, C.: Student Dropout Prediction in Higher Education: A Systematic Review of Machine Learning Methods and Risk Factors. Journal of Information Technology Education: Research (2026). doi:10.28945/5776
- [12] M. Nagy and R. Molontay, "Interpretable Dropout Prediction: Towards XAI-Based Personalized Intervention," International Journal of Artificial Intelligence in Education, vol. 34, pp. 274–300, 2024, doi: 10.1007/s40593-023-00331-8.
- [13] V. Dermol and V. Skrbinjek, "Predicting student dropout in higher education using Random Forest and survival analysis," Discover Education, 2026, doi: 10.1007/s44217-026-01710-8.
- [14] Prediction of student dropout using Artificial Intelligence algorithms and machine learning algorithms - ScienceDirect [https://www.sciencedirect.com/science/article/pii/S1877050924034161?cf\\_chl\\_tk=0d69jCpSP2IYLFbKEG89BRUFzB604hz31NgkUV91A-1787152266-1.0.1.1-66LWAGDTRjKsRzLTDOiPETVG.LU\\_d7XqB32MH2kzGIA](https://www.sciencedirect.com/science/article/pii/S1877050924034161?cf_chl_tk=0d69jCpSP2IYLFbKEG89BRUFzB604hz31NgkUV91A-1787152266-1.0.1.1-66LWAGDTRjKsRzLTDOiPETVG.LU_d7XqB32MH2kzGIA)
- [15] A chart showing categories of student dropout factors - [https://www.google.com/imgres?imgurl=https://charlatan.ca/wp-content/uploads/2026/06/Screenshot-2026-06-12-003208.png&tbid=Q0bPpD5Uj\\_88SM&vet=1&imgrefurl=https://charlatan.ca/educated-in-the-age-of-ai-the-class-of-2026-prepares-to-graduate/&docid=UVhPhH9lt3HXyM&w=1015&h=482&hl=en-US&source=sh/x/im/m1/4&kgs=0c07f3c35b0af171&shem=epsd1\\_nisbtsa2.nisbtsal.rimspwouoe&utm\\_source=epsd1.nisbtsa2.nisbtsal.rimspwouoe.sh/x/im/m1/4](https://www.google.com/imgres?imgurl=https://charlatan.ca/wp-content/uploads/2026/06/Screenshot-2026-06-12-003208.png&tbid=Q0bPpD5Uj_88SM&vet=1&imgrefurl=https://charlatan.ca/educated-in-the-age-of-ai-the-class-of-2026-prepares-to-graduate/&docid=UVhPhH9lt3HXyM&w=1015&h=482&hl=en-US&source=sh/x/im/m1/4&kgs=0c07f3c35b0af171&shem=epsd1_nisbtsa2.nisbtsal.rimspwouoe&utm_source=epsd1.nisbtsa2.nisbtsal.rimspwouoe.sh/x/im/m1/4)
- [16] Flowchart of the student dropout prediction system – [https://www.google.com/imgres?imgurl=x-raw-image:///357c7ff715ac12d816299a01d99500f5f86eacaf5360d5bf5a525039b1c40d8&tbid=2BKsFfd\\_fDX-TM&vet=1&imgrefurl=https://zenodo.org/records/13745100/files/An%2520AI-Driven%2520Predictive%2520Formatted%2520Paper.pdf?download%3DI&docid=OvDzX9SuOC8kLM&w=386&h=560&hl=en-US&source=sh/x/im/m1/4&kgs=b5c943899f9a80a3&shem=epsd1\\_nisbtsa2.nisbtsal.rimspwouoe](https://www.google.com/imgres?imgurl=x-raw-image:///357c7ff715ac12d816299a01d99500f5f86eacaf5360d5bf5a525039b1c40d8&tbid=2BKsFfd_fDX-TM&vet=1&imgrefurl=https://zenodo.org/records/13745100/files/An%2520AI-Driven%2520Predictive%2520Formatted%2520Paper.pdf?download%3DI&docid=OvDzX9SuOC8kLM&w=386&h=560&hl=en-US&source=sh/x/im/m1/4&kgs=b5c943899f9a80a3&shem=epsd1_nisbtsa2.nisbtsal.rimspwouoe)