International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.061

e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 382-391

Smart Drop: A Deep Learning Framework for Predicting College Dropouts

Deepa Priya V[1], Gabriel Anto Joshua S[2], Jayashree T[3], Sandhiya K[4], Srideepalakshmi S[5], Harini U[6]

[1,3,4,5,6] Department of Information Technology, Kamaraj College of Engineering and Technology, Virudhunagar, Tamilnadu, India.

[2] Department of Computer Science and Engineering, Kamaraj College of Engineering and Technology, Virudhunagar, Tamilnadu, India.

Email ID: *deepapriyakcet@gmail.com*

Abstract

Educational Data Mining (EDM) plays a key role in improving modern learning by using advanced techniques. This study focuses on predicting college dropout rates with a deep learning framework. By analysing student data, our model helps identify at-risk students early, allowing timely support. The dataset contains student details such as academic performance, socioeconomic status, and engagement levels. Using these details, our framework applies deep learning methods to detect early signs of dropout. This helps universities take quick action to support students and improve their learning experience. Our study checked how well different models predicted dropout rates. The K-Nearest Neighbours (KNN) model had 93% accuracy, while the Recurrent Neural Network (RNN) had 95% accuracy. When KNN and RNN were combined, accuracy dropped to 87%. However, combining KNN and Convolutional Neural Network (CNN) improved accuracy to 97%. The RNN and CNN combination scored 96% accuracy, and the standalone CNN model also reached 97% accuracy. The best-performing method in this study is the KNN and CNN combination, which achieved the highest accuracy of 97%. Even though the CNN model alone also reached 97%, combining it with KNN helps capture more patterns in the data. This approach improves student retention by identifying and helping struggling students early. Our findings show that deep learning can effectively predict dropout rates, allowing educational institutions to provide better student support and improve success rates.

Keywords: *KNN, RNN, CNN, Deep Learning, Machine Learning, Neural Networks*

1. Introduction

In recent years, technology has changed the way students learn. A major problem in education is that many students leave college before finishing. This not only affects students but also makes academic programs less effective. To fix this, new ways using technology and data are needed. Educational Data Mining (EDM) helps collect useful information from student records. By looking at student performance, participation, and background, EDM helps find reasons why students drop out. Deep learning is also becoming popular because it can find patterns and make good predictions. This paper presents a novel deep learning framework designed to predict college dropout rates and enhance the overall learning experience for students. By harnessing advanced neural network architectures, including K-nearest neighbours (KNN) and recurrent neural networks (RNN), this framework scrutinizes diverse student data to identify early indicators of potential dropout. Through timely interventions, institutions can proactively support at-risk students, thereby mitigating dropout rates and fostering a more inclusive learning environment. The efficacy of the proposed framework is underscored by the impressive accuracy rates achieved by the KNN and RNN models—94% and 92%, respectively. These results demonstrate the potential of deep learning to revolutionize academic support by providing actionable insights and empowering institutions to intervene effectively. In the following sections, we delve into the methodology, results, and implications

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.061
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 382-391

of this pioneering approach, emphasizing its significance in advancing educational research and practice. By embracing the convergence of data science and education, we aim to catalyse positive change and elevate the educational experience for all stakeholders. [1-5]

1.1 Need and Motivation

1.1.1 Improved Student Retention

By accurately predicting dropout rates, institutions can implement timely interventions to support at-risk students, thereby increasing student retention rates. This leads to higher graduation rates and a more successful student body.

1.1.2 Enhanced Learning Experience

Proactive interventions not only prevent dropout but also contribute to a more supportive and engaging learning environment. Students receive personalized support tailored to their needs, leading to improved academic outcomes and overall satisfaction with their educational experience.

1.1.3 Resource Optimization

Identifying at-risk students early allows institutions to allocate resources more efficiently. Instead of waiting until students are in crisis, resources can be directed towards prevention and proactive support measures, optimizing the use of time and resources.

1.1.4 Data-Driven Decision Making

The research leverages educational data mining and deep learning techniques to extract insights from large datasets. This promotes data-driven decision-making in academic support services, enabling institutions to base interventions on empirical evidence rather than intuition alone.

1.1.5 Long-term Impact

By addressing the root causes of dropout rates and providing effective support mechanisms, the research has the potential to have a lasting impact on students' lives. Graduating from college opens doors to higher-paying jobs, better career opportunities, and improved socio-economic outcomes for individuals and communities. Overall, the research contributes to the advancement of educational practices by harnessing technology to promote student success and create more inclusive learning environments. support at-risk students, thereby mitigating dropout rates and fostering

2. Literature Review

The problem of college dropouts is a big concern for teachers, researchers, and policymakers. Studies show that about 40% of college students in the United States leave before finishing their degrees. This issue is more common for students from low-income families (NCES, 2020). Leaving college early can make it harder to find good jobs and can lower future income. It also affects society by slowing down economic growth and making social inequality worse (Belfield & Bailey, 2011). To fix this problem, researchers are using artificial intelligence (AI) and machine learning (ML) to find students who might drop out. These technologies look at student data to find signs of dropout risk. The goal is to create systems that help students stay in college. [1] In the study on predicting student dropout based on machine learning and deep learning, various models were explored, including Neuronal Networks (NN), Decision Tree (DT), Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Gradient Boosted Tree, Naïve Bayes (NB), Generalized Linear Model (LightGBM), K-NN, Adaptive Boost (AdaBoost), XGBoost, CatBoost, and Bayesian networks (BNs). The models were evaluated using metrics such as accuracy, with some achieving high accuracies ranging from 66.5% to 99%. Among these, the combination of Neuronal network (NN), Decision Tree (DT), and Bayesian networks (BNs) achieved an accuracy of 84.8%. The study provides a comprehensive overview of the performance of various machine learning and deep learning models in predicting student dropout. [2] In the study on predicting student dropout with minimal information, several models were examined, including Logistic Regression (LR), Linear Discriminant Analysis (LDA), Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN). The models achieved accuracies ranging from 76.9% to 97.2%, with SVM achieving the highest accuracy of 97.2%. These results suggest that even with minimal information, machine learning models can effectively predict student dropout, with SVM performing particularly well in

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.061
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 382-391

this context. [3] In the International Journal of Data Mining & Knowledge Management Process (IJDKP), a study was conducted using Decision Tree and Naïve Bayes algorithm to predict student dropout. The Decision Tree model achieved an accuracy of 98.14%, while the Naïve Bayes algorithm achieved an accuracy of 96.98%. These results indicate the effectiveness of both models in predicting student dropout rates. [4] In the MDPI study, several machine learning models, including Artificial Neural Networks, Gradient Boosted Tree, Support Vector Machine, Decision Tree, Random Forest, and XGBoost, were employed to predict student dropout rates. The models achieved varying levels of accuracy, with Decision Tree and Random Forest performing notably well, achieving accuracies of 0.89 and 0.93, respectively. Support Vector Machine also performed reasonably well with an accuracy of 0.78, while Artificial Neural Networks, Gradient Boosted Tree, and XGBoost achieved accuracies of 0.75, 0.74, and 0.90, respectively. These results suggest that Random Forest and Decision Tree are effective models for predicting student dropout rates. [5] In the study on university student dropout prediction using pretrained language models, various machine learning models were applied, including Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, Multilayer Perceptron (MLP), and MSNF. These models were evaluated based on their accuracy, with MSNF achieving the highest accuracy of 0.877, followed by SVM with 0.847, Random Forest with 0.829, Decision Tree with 0.821, MLP with 0.896, and Logistic Regression with 0.766. These results suggest that pretrained language models can be effective in predicting student dropout rates, with MLP performing the best among the models tested. [6] In the study on the analysis of first-year university student dropout through machine learning models, a comparison was made between various universities using different models. The models included Random model, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest, Gradient Boosting, Naive Bayes, Logistic Regression, and Neural Network. The performance of these models was evaluated based on their accuracy, with Neural Network achieving the highest accuracy of 0.69 ± 0.02, followed by Logistic Regression and Gradient Boosting with accuracies of 0.66 ± 0.01 and 0.69 ± 0.02, respectively. Naive Bayes, Random Forest, and Decision Tree also performed well with accuracies of 0.69 ± 0.02, 0.68 ± 0.03, and 0.65 ± 0.02, respectively. SVM and KNN had accuracies of 0.62 ± 0.02 and 0.62 ± 0.02, while the Random model had the lowest accuracy of 0.51 ± 0.02. These results indicate that Neural Network, Logistic Regression, and Gradient Boosting are effective models for predicting first-year university student dropout, with Neural Network performing the best among the models tested. [7] In the study on using neural networks to predict dropout at universities, two models were employed: Perceptron Multilayer and Radial Basis Function. These models achieved high accuracies, with Perceptron Multilayer reaching 98.6% and Radial Basis Function achieving 98.1%. These results suggest that neural networks, particularly the Perceptron Multilayer model, can be highly effective in predicting dropout rates at universities. [8] In the study on predicting student dropout in university courses using different machine learning techniques, several models were evaluated for their effectiveness. Naïve Bayes (NB) achieved an accuracy of 0.77, while Random Forest (RF) and Logistic Regression (LR) both performed well with accuracies of 0.93. Support Vector Machines (SVM) and Decision Tree (DT) also showed good performance with accuracies of 0.92 and 0.90, respectively. Neural Network (NN) achieved an accuracy of 0.88. These results suggest that Random Forest, Logistic Regression, SVM, and DT are effective models for predicting student dropout in university courses, while Naïve Bayes and Neural Network are slightly less accurate in this context. [9] In the study on MOOC dropout prediction using the FWTS-CNN model based on Fused Feature Weighting and Time Series, several machine learning models were compared, including Logistic Regression (LR), Naïve Bayes (NB), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), Convolutional Neural

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.061
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 382-391

Network (CNN), and the proposed FWTS-CNN model. Among these models, FWTS-CNN achieved the highest accuracy of 0.871, followed by CNN with an accuracy of 0.853. SVM also performed well with an accuracy of 0.681, while LR, NB, RF, and DT had accuracies of 0.671, 0.673, 0.658, and 0.622, respectively. These results suggest that the FWTS-CNN model is effective for predicting MOOC dropout rates and outperforms traditional machine learning models in this context. [10] In the study on student dropout prediction using machine learning techniques, several models were evaluated for their effectiveness. Naïve Bayes (NB) achieved an accuracy of 0.92, while Neural Networks (NN), Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), and Logistic Regression (LR) all performed equally well with accuracies of 0.93. These results indicate that all these models are effective for predicting student dropout rates, with NN, RF, SVM, DT, and LR showing particularly high accuracies. [11] In the study on student dropout prediction, Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and Random Forest (RF) were evaluated for their effectiveness. LDA and SVM both achieved an accuracy of 0.62, while RF had a lower accuracy of 0.56. These results suggest that LDA and SVM are more effective models for predicting student dropout compared to Random Forest. [12] In the study on predicting student dropout in university classes using a two-layer ensemble machine learning approach, several models were utilized, including Feed Forward Neural Network, Random Forest, Gradient Boosting, Extreme Gradient Boosting, and Stacking ensemble. The Stacking ensemble model achieved the highest accuracy of 92.18%, outperforming the other models. Random Forest and Extreme Gradient Boosting also performed well with accuracies of 91.66%, while Gradient Boosting achieved an accuracy of 86.66%. Feed Forward Neural Network had the lowest accuracy among the models, with 76.67%. These results indicate that the Stacking ensemble approach is effective for predicting student dropout rates in university classes. [13] In the study on the global challenges of student dropout and the development of a prediction model using machine learning algorithms on higher education datasets, several models were evaluated for their effectiveness. Decision Tree and Random Forest both achieved accuracies of 0.93, while Logistic Regression performed slightly better with an accuracy of 0.94. Naïve Bayes had a lower accuracy of 0.88. These results suggest that Logistic Regression is the most effective model among those tested for predicting student dropout in higher education settings. [14] In the study on predicting student dropout in online classes using a DeepFM-based predictive model, the DeepFM model achieved an impressive accuracy of 99%. This suggests that the DeepFM model is highly effective for predicting student dropout in the context of online classes. [15] In the study on a time-aware approach for MOOC dropout prediction based on rule induction and sequential three-way decisions, the models used were Rule Induction, Machine Learning Classifiers with enriched feature sets, and Sequential Three-Way Decision. However, the specific accuracy or performance metrics for these models were not provided (NA). [16] In the study comparing automated machine learning against a shelf pattern classifier in predicting university dropout, SVM achieved an accuracy of 0.87. The specific accuracies for Automated Machine Learning and the Off-the-Shelf pattern-based classifier were not provided. The above-mentioned survey provides proof of the neural network algorithm, indicating that machine learning techniques yield the best results with a high accuracy rate. Thus, one machine learning method is also used in our proposed system. It is known as a K-nearest neighbor algorithm. [6-10]

3. Proposed Work

Our methodology for predicting student dropout rates involves a systematic approach, as illustrated in the following flowchart. We emphasize a clear distinction between the training and testing phases of our models. In the training phase, we utilize both traditional machine learning algorithms and advanced deep learning techniques to train our models on a comprehensive dataset comprising student demographics, academic performance

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.061
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 382-391

metrics, and engagement indicators. Following the training phase, we rigorously evaluate the performance of our models using established metrics such as accuracy, precision, recall, and F1 score. This evaluation process ensures that our models are robust and capable of accurately predicting student dropout rates in real-world scenarios. By adopting this methodology, we aim to provide a reliable and effective framework for predicting student dropout rates, which can inform targeted interventions and support mechanisms to improve student retention and academic success.

3.1 Data Source and Description

In the proposed system architecture, the system is provided with the dataset which is taken from Kaggle [14] which is an Open-source platform.NThis dataset provides a comprehensive view of students enrolled in various undergraduate degrees offered at a higher education institution. It has 4425 records and 35 columns. It includes demographic data, socioeconomic factors, and academic performance information that can be used to analyze the possible predictors of student dropout and academic success. This dataset contains multiple disjoint databases consisting of relevant information available at the time of enrolment, such as application mode, marital status, course chosen, and more. Additionally, this data can be used to estimate overall student performance at the end of each semester by assessing curricular units credited/enrolled/evaluated/approved as well as their respective grades. Finally, we have the unemployment rate, inflation rate, and GDP from the region which can help us further understand how economic factors play into student dropout rates or academic success outcomes. This powerful analysis tool will provide valuable insight into what motivates students to stay in school or abandon their studies for a wide range of disciplines such as agronomy, design, education nursing journalism management social service, or technologies. Figure 1 shows The Work Flow of Predicting Students Dropout Rate predicting student dropout with minimal information, several models were examined, including Logistic Regression (LR), Linear Discriminant [11-15]

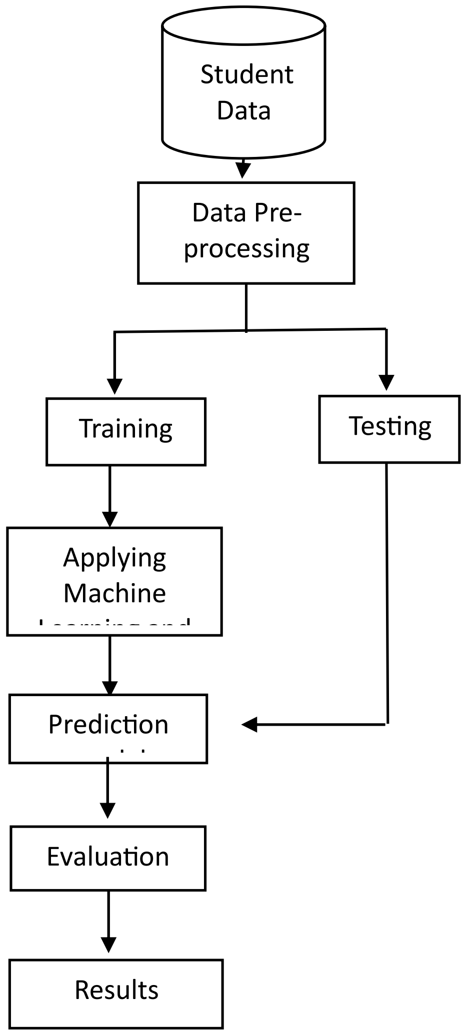

Figure 1 The Work Flow of Predicting Students Dropout Rate

3.2 Data Pre-Processing

Data Preprocessing is a crucial step in the data analysis pipeline where raw data is transformed into a clean, understandable format that is suitable for analysis. It involves a series of operations such as cleaning, transforming, and organizing data to make it suitable for further analysis by machine learning algorithms or statistical methods. There are several Preprocessing methods used for this proposed system.

3.2.1 Handling Missing Values

- Dropna () – It is used to remove any rows from the dataset where there are missing values(NaNs). It effectively drops those rows

e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 382-391

from the dataset, allowing you to work with clean data that doesn't contain missing values.

- Categorical_features contains the names of categorical columns in the dataset (columns with data type object).

3.2.2 Scaling Numerical Features and Categorical Features

- X contains the features, where the 'Daytime/even attendance' column is dropped to exclude it from the features.
- Y contains the target variable, which is the 'Daytime/even attendance' column.
- Numerical_features contain the names of numerical columns in the dataset (columns with data types float64 or int 64).

4. Methodology

The methodology employed in this study encompasses a comprehensive framework that leverages both deep learning and traditional machine learning techniques to address the research objectives effectively. With the overarching goal of predicting college dropout rates and enhancing learning experiences, a multifaceted approach integrating various models and algorithms is adopted.

4.1 KNN (K-Nearest Neighbor)

K-Nearest Neighbours (KNN) stands as a beacon of simplicity in the vast landscape of machine learning algorithms. Its charm lies in its straightforward approach: no complex assumptions, no intricate parameters. Instead, KNN operates on the principle of proximity, making it a versatile choice for classification and regression tasks where discerning patterns amidst data is paramount. KNN is useful for solving classification problems. It allows us to make predictions based on the majority class among the `k` nearest neighbours. By using the KNN algorithm we can identify the nearest neighbors for each data point in the test set and predict their labels accordingly. predicting student dropout with minimal information, several models were examined, including Logistic Regression (LR), Linear Discriminant Analysis (LDA), Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN). The basic KNN architecture is represented in Figure 2.

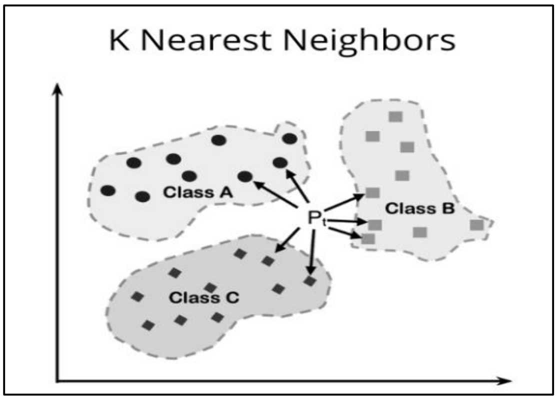

Figure 2 Basic KNN Architecture

4.2 RNN (Recurrent Neural Network)

A recurrent neural network (RNN) is a type of artificial neural network that uses sequential data or time series data. These deep learning algorithms are commonly used for ordinal or temporal problems, such as language translation, natural language processing (NLP), speech recognition, and image captioning; they are incorporated into popular applications such as Siri, voice search, and Google Translate. The basic RNN architecture is represented in Figure 3.

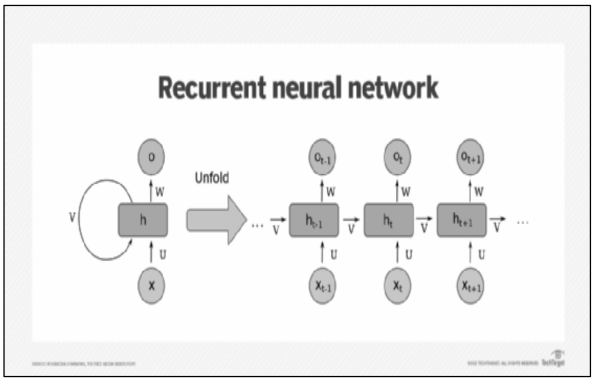

Figure 3 Basic RNN Architecture

The RNN captures dependencies between successive data points and makes predictions based on the entire sequence. In classification tasks, RNNs can identify patterns in the sequential data and predict the corresponding class labels. This makes RNNs particularly useful for tasks such as sentiment analysis in text data or speech recognition in audio data, where the order of the data points is essential for accurate predictions.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.061
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 382-391

4.3 CNN (Convolutional Neural Network)

A Convolutional Neural Network or CNN is a type of artificial neural network, which is widely used for image/ object recognition and classification. Deep Learning thus recognizes objects in an image by using a CNN. The basic CNN architecture is represented in Figure 4.

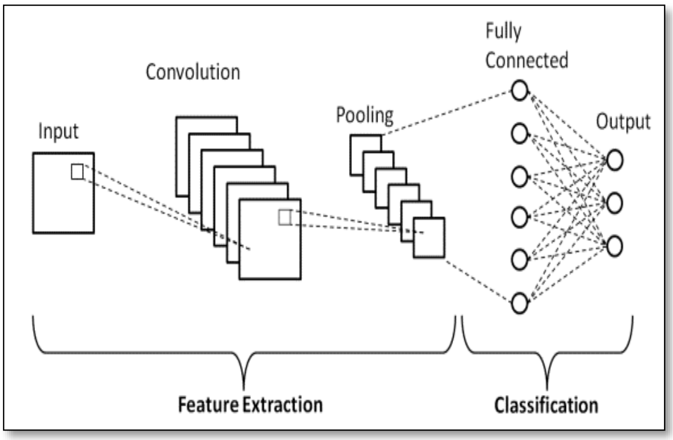

Figure 4 Basic CNN Architecture

The basic CNN comprises three main layers: the convolutional layer, the pooling layer, and the fully connected layer.

5. Model Validation

There are a variety of measures for various algorithms and these measures have been developed to evaluate very different things. So it should be the criteria for evaluation of various proposed methods. False Positive (FP), False Negative(FN), True Positive(TP), True Negative(TN), and the relation between them are quantities usually adopted by predicting students' dropout rates to compare the accuracy of different approaches. The definitions of the mentioned parameters are presented below:

Confusion Matrix: This offers more information about a predictive model's performance as well as which classes are being forecasted correctly, which erroneously, and what kinds of mistakes are being produced. Figure 5 shows Confusion Matrix. future income. It also affects society by slowing down economic growth and making social inequality worse (Belfield & Bailey, 2011). To fix this problem, researchers are using.

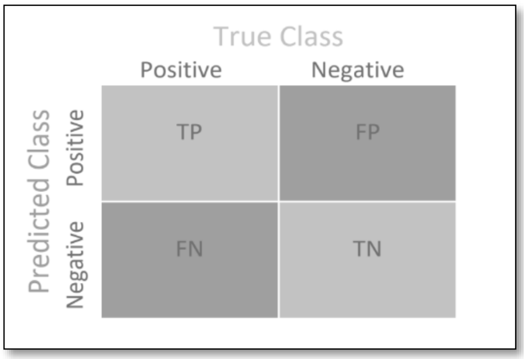

Figure 5 Confusion Matrix

Accuracy: The proportion of accurately identified instances is known as accuracy. One of the most used measures for measuring categorization performance is this one. Number of accurate forecasts / Total number of predictions equals accuracy. Alternatively, the accuracy of binary classification models is defined as follows:

Accuracy = TP+TN / (TP+TN+FP+FN)

Precision and Recall: Precision is the quantity of positively identified occurrences that are fraudulent or classed as positive examples.

Precision = TP/(TP+FP)

A measure known as recall counts the number of accurate positive predictions among all possible positive predictions. Recall shows missed positive predictions, in contrast to precision, which only comments on the accurate positive predictions out of all positive predictions. The calculation of recall involves dividing the total number of true positives and false negatives by the number of true positives.

Recall = TP / (TP + FN)

F1 Score: Precision and Recall are weighted averages that make up the F1 Score. As such, this score accounts for both false positives and false negatives.

F1 Score = 2*(Recall * Precision) / (Recall + Precision).

TP represents True Positive

TN represents True Negative

FP represents False Positive

FN represents False Negative

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.061
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 382-391

6. Result and Analysis

Based on the evaluation results obtained from this proposed system, we observed distinct performance trends across different models and combinations. Firstly, the K-Nearest Neighbors (KNN) model demonstrated solid performance with an accuracy of 93% using 5 neighbors. Although it achieved a slightly lower accuracy compared to other models, it showcased high recall, precision, and F1-score, indicating its effectiveness in identifying true positives. Figure 6 shows Metrics of KNN Algorithm.

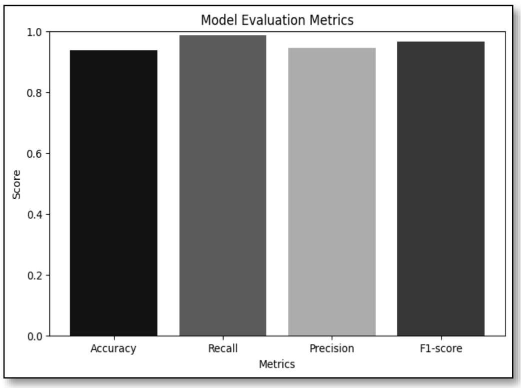

Figure 6 Metrics of KNN Algorithm

Next, the Recurrent Neural Network (RNN) model exhibited improved accuracy, reaching 95% after 35 epochs. It displayed consistently high recall, precision, and F1-score, showcasing its ability to capture temporal dependencies and patterns in the data. Figure 7 shows Metrics of the RNN Algorithm.

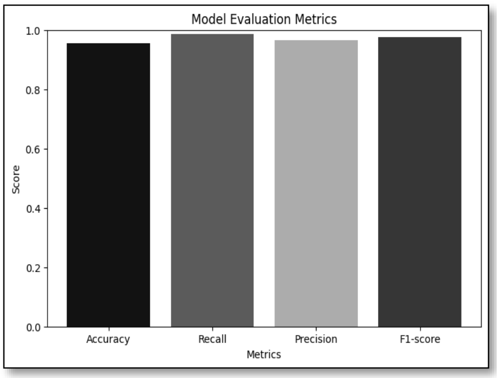

Figure 7 Metrics of the RNN Algorithm

The Convolutional Neural Network (CNN) emerged as the top-performing model, achieving the highest accuracy of 97% after 35 epochs. With superior recall, precision, and F1-score, CNN demonstrated its capability to extract spatial features and relationships, leading to enhanced predictive performance. Figure 8 shows Metrics of CNN Algorithm

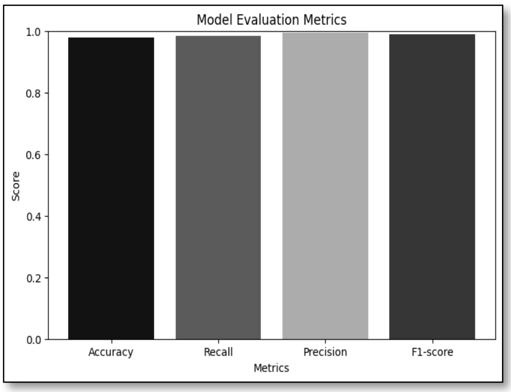

Figure 8 Metrics of CNN Algorithm

The above Figures 6,7 and 8 represent the metrics of accuracy, recall, precision, and f1-score of the KNN, RNN, and CNN algorithms. When combining models, the fusion of KNN & RNN yielded a lower accuracy of 87%, yet maintained high recall, precision, and F1-score. On the other hand, both KNN & CNN and RNN & CNN combinations resulted in competitive accuracies of 97% and 96%, respectively, along with exceptional performance across all evaluation metrics.

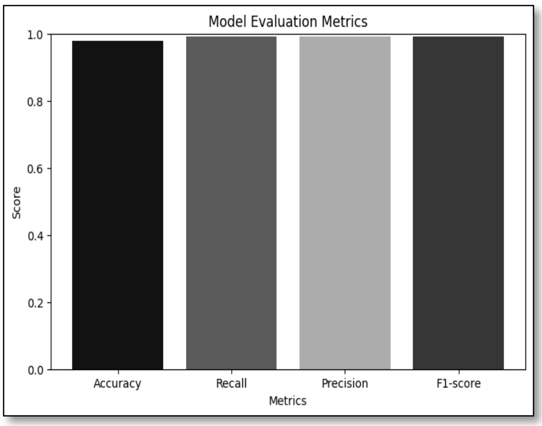

Figure 9 Metrics of KNN & CNN Combination

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.061
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 382-391

Figure 9 shows the metrics of the accuracy, recall, precision, and f1-score of the best model of the KNN and CNN model. In this comparative study, the KNN & CNN combination method yielded higher accuracy. So in this proposed system, that model is the best one. [16-17]

Conclusion

In conclusion, the proposed work focused on predicting student dropout rates using a range of machine learning models, including K-Nearest Neighbors (KNN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN). Through meticulous data collection, Preprocessing, and model training, we aimed to uncover insightful patterns and factors contributing to dropout risks, thereby facilitating proactive interventions to enhance student retention and academic success. Firstly, comprehensive student data encompassing demographics, academic performance, and attendance records was collected and preprocessed to ensure its suitability for modeling. Subsequently, tailored Preprocessing steps were applied to accommodate the specific requirements of each model, such as encoding categorical variables or resizing images. Upon model selection, rigorous training and hyper parameter tuning were conducted to optimize performance. The effectiveness of each model was evaluated using established metrics like accuracy, precision, recall, and F1-score on separate testing datasets. By comparing the performance across different models, we sought to identify the most reliable approach for predicting dropout rates accurately. In essence, the proposed work represents a proactive approach to addressing student dropout rates through the integration of advanced machine learning techniques. By harnessing predictive analytics, educational institutions can take primitive measures to support at-risk students, thereby fostering a more inclusive and supportive academic environment. In conclusion, the evaluation results highlight the effectiveness of CNN in predicting student dropout rates, outperforming both KNN and RNN individually. The above figure 10 shows that combining CNN with either KNN or RNN further enhances predictive accuracy and model robustness, suggesting the potential benefits of ensemble methods in addressing complex classification tasks. These findings underscore the importance of leveraging diverse modeling techniques and ensemble strategies to achieve optimal performance in educational predictive analytics. Further, the future work of this proposed system is to Implement real-time monitoring systems that leverage predictive models to identify students at risk of dropout early on. These systems could provide timely interventions and personalized support to students, improving retention rates and academic outcomes.

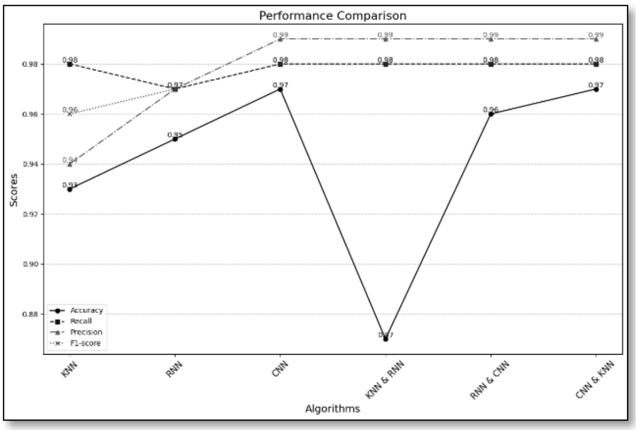

Figure 10 Comparison Performance

References

[1]. Hyun-Sik Won, et al. "University Student Dropout Prediction Using Pretrained Language Models" International Journal for Research in Applied Science &amp on 13 June 2023 https://www.mdpi.com/2076-3417/13/12/7073

[2]. Diego Opazo, et al. "Analysis of First-Year University Student Dropout through Machine Learning Models: A Comparison between Universities" Mathematics on 15 October 2021 https://www.mdpi.com/2227-7390/9/20/2599

[3]. Mayra Alban , et al.“Neural Networks to Predict Dropout at the Universities” International Journal of Machine Learning and Computing in 2, April 2019 https:// www.i jmlc. org/ vol 9/779-ML0074.pdf

[4]. Janka Kabathova, et al. “Towards Predicting Student’s Dropout in University Courses Using Different Machine Learning Techniques”

International Journal for Research in Applied Science &amp on 1 April 2021 https://www.mdpi.com/2076-3417/11/7/3130
[5]. YAFENG ZHENG ,et al " MOOC Dropout Prediction Using FWTS-CNN Model Based on Fused Feature Weighting and Time Series" IEEE in December 16, 202 https://pdfs .semanticscholar.org/a328/6f0ed69103e473ef09 977ce9b28683418bc5.pdf
[6]. Francesca Del Bonifro, et al."Student Dropout Prediction". SPRINGER LINK, 12163, (2020):129–140. https:// link.springer .com/chapter/10.1007/ 978-3-030-52237-7_11#Sec5
[7]. Haarika Dasi1, et al. "Student Dropout Prediction Using Machine Learning Techniques". International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING,(2020): ISSN:2147-6799. https://ijisae.org/index.php/IJISAE/article/view/2276/859
[8]. Jovial Niyogisubizo a, et al." Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization." Science Direct, 28-Mar-22 " Volume 3, 2022, 100066. https://www.sciencedirect.com/science/article/pii/S2666920X22000212?ref=pdf_download&fr=RR-2&rr=85f748f8e8ed7ecd
[9]. Meseret Yihun Amare, et al. "Global challenges of students dropout: A prediction model development using machine learning algorithms on higher education datasets ." SHS Web of Conferences, 129, 09001 (2021). https://www.shsconferences.org/articles/shsconf/pdf/2021/40/shsconf_glob2021_09001.pdf
[10]. Daniel Andrade Girón, et al. "Predicting Student Dropout based on Machine Learning and Deep Learning" EAI Endorsed Transactions on Scalable Information Systems in 18 July 2023.https://publications.eai.eu/index.php/sis/article/view/3586
[11]. Nassim Mouchantaf, Maroun Chamoun," Predicting Student Dropout with Minimal Information" Iraqi Journal of Science on October 30, 2023.https://www.iasj.net/iasj/article/291323
[12]. Ghadeer S, et al." Data Mining In Higher Education: University Student Dropout" International Journal of Data Mining & Knowledge Management Process (IJDKP) on January 1, 2020. https:// www.semanticscholar.org/paper/Data-mining-in-highereducation%3A-University-Student-Abu-Oda-El Halees/ f1b5f4320d381 cdb9807a03a81938eda2d166284
[13]. Zihan Song, et al." All-Year Dropout Prediction Modelling and Analysis for University Students." MDPI on 14 January 2023. https://www.mdpi.com/2076-3417/13/2/1143
[14]. Dataset: Predict Student Dropout rates and academic success https://www.kaggle.com/datasets/ thedevastator/higher-education-predictors-of-student-retention
[15]. Nuha Mohammed Alruwais, "Deep FM-Based Predictive Model for Student Dropout in Online Classes" Research Article, 11,2023 https://ieeexplore.ieee.org/document/10239344
[16]. Carlo Blundo, et al." A Time-Aware Approach for MOOC Dropout Prediction Based on Rule Induction and Sequential Three-Way Decisions" Research Article,11,2023 https://www.researchgate.net/publication/374611309_A_time aware_approach_for_MOOC_dropout_prediction_based_on_Rule_Induction_and_Sequential_Three-Way_Decisions
[17]. Leonardo Canete-Sifuentes, et al "Comparing Automated Machine Learning against an Off the shelf Pattern Based Classifier in a Class imbalance problem: Predicting University Dropout", Preparation of Papers for IEEE TRANSACTIONS and JOURNALS, 11,2023