International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4274-4283
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0625

Addressing Gaps in Scholarship Recommendations: A Review Focusing on State

Aneena Alphonse Sunny[1], Vishnu Mohan C[2]
[1]*Scholar, Department of Computer Science, Sacred Heart College (Autonomous), Thevara, Kerala, India.*
[2]*Assistant Professor, Department of Computer Science, Sacred Heart College (Autonomous), Thevara, Kerala,India.*
Emails: *neenaalphonsesunny@gmail.com[1], vishnumohan@shcollege.ac.in[2]*

Abstract
Existing scholarship recommendation systems typically provide only acceptance or rejection outcomes, without offering explanations. This limitation is evident in Kerala, where students often struggle to interpret the numerous complex rules governing scholarships. This review examines existing recommendation approaches, focusing on methods that explain their decisions for merit-based and minority-based schemes. The review evaluates techniques such as rule-based systems, collaborative filtering, content-based filtering, and decision trees, assessing their strengths and weaknesses in providing feedback, personalizing recommendations, and handling challenges such as dispersed information and intricate eligibility criteria. A key observation is that current systems rarely explain rejection reasons or propose alternative opportunities tailored to specific regions like Kerala. The findings of this review aim to guide the development of a scholarship recommendation system using Python's Scikit-learn, Pandas, and Streamlit. This system will apply an XGBoost-based decision tree ensemble to recommend Kerala-specific merit and minority scholarships, clearly explain rejections, and suggest alternative schemes.

Keywords: *Scholarship Eligibility Prediction, Explainable AI, Educational Data Mining, Fairness In Machine Learning, XGBoost, Ensemble Learning, Recommender Systems, Kerala Scholarships, Minority Scholarships.*

1. Introduction
Scholarships are important in giving deserving students a fair chance to continue their education. They help reduce financial barriers and support educational equality. However, in many places, scholarship systems only tell applicants if they are 'accepted' or 'rejected' without explanation.This lack of explanation makes it hard for students to know what went wrong or how they can improve their chances next time.In Kerala, this problem is even bigger because there are many different merit-based and minority-based scholarship schemes where each has its own rules, such as required marks, income limits, and community reservations. These rules can be complex, making them difficult to understand. Students often have difficulty finding information, checking eligibility, and deciding which scholarships to apply for. As a result, some qualified students miss opportunities, while others waste time applying for scholarships they cannot get. To address these challenges, there is a need for a new methodology that combines machine learning, explainable AI (XAI), and rule-based verification to create a transparent and fair recommendation system. The proposed model leverages the XGBoost algorithm for accurate prediction of eligibility, integrated with SHAP-based explainability to communicate clear, human-understandable reasons for every decision. Furthermore, the system introduces a recommendation component that suggests alternative scholarships when a student is ineligible, ensuring inclusivity and continuous opportunity. The goal is to use these ideas to design a scholarship system for Kerala that is accurate, transparent, fair, and easy for students to use.

2. Literature Review: Building An Explainable and Fair System for Kerala

1. The Specific Context of Kerala's Scholarships
Kerala has a diverse set of scholarship schemes, administered by various state and central government departments, targeting different categories such as

merit, minority status, and financial need. While these programs aim to improve accessibility to higher education, the lack of a unified and transparent platform creates barriers for students. Many applicants face rejections without understanding the reasons, which undermines trust in the system.

- **C.H. Muhammedkoya Scholarship:** This scholarship is for talented minority girls from specific communities like Muslim, Latin, and Converted Christian. To get it, a student must have at least 50% marks and a family income less than 8 lakh per year. The scholarship gives different amounts of money for different courses, plus extra money for hostel costs. This shows that a good system must check many different things such as if they get extra money for things like a hostel.
- **State Merit Scholarship (SMS):** This is for first-year students in government and aided colleges. You need at least 50%, but the family income limit is very strict: only 1 lakh per year. This low income limit means your eligibility checker needs to be very accurate to make sure only the right students are chosen.
- **Kerala State Higher Education Council Scholarship (HECS):** This scholarship is for first-year degree students. The marks you need to get in depend on your category. For example, disabled students need 45%, but general students need 60-75%. Also, the money you get changes each year. This means a good recommendation system must be able to keep track of these changing rules.
- These examples show that a Kerala-specific scholarship system needs to deal with many different, changing, and sometimes overlapping rules. A clear, single platform could make the process easier to manage, give students better access to information, and help them understand the rules more easily.

2. The Case for Interpretable Models

A big challenge in building scholarship recommendation systems is to go beyond “yes” or “no” answers. Students, especially in competitive scholarships, want to know why they were accepted or rejected. Interpretable models, like Decision Trees and rule-based systems, are good at making predictions while also clearly showing the reasons behind them. Roslan and Chen in their systematic review analyzed various educational data mining (EDM) techniques for predicting student performance.[1] Their findings identified Decision Trees as the most widely adopted and effective approach. The authors highlight a key advantage of this model: its ability to balance high predictive accuracy with interpretability. This transparency is crucial for applications like scholarship eligibility, where understanding the decision-making process is essential for building trust. A central theme of their work is the importance of addressing fairness and bias. They emphasize the need to actively identify and mitigate biases in predictive models to prevent discrimination against specific demographic groups. This insight is highly relevant to scholarship systems, where it is vital to ensure equitable access. The research of Roslan and Chen provides a strong foundation for the development of an explainable, fair, and reliable eligibility system, particularly in complex environments such as Kerala’s merit and minority scholarship schemes. Khan et al. built a decision tree to monitor students’ progress in the middle of a semester [2]. It could detect students who are likely to receive low grades so help could be given early. In scholarships, this idea can be used to check if students are still eligible during the year, especially for scholarships that continue for several years. The transparent nature of the Decision Tree model would ensure that any decision to review a student’s funding is based on clear, understandable criteria, which is essential for maintaining fairness and trust in the scholarship allocation process. Bhegade and Shinde used Decision Trees along with a technique called FP-growth to find patterns in the profiles of successful students.[3] This gave both accurate predictions and useful insights, such as common traits of award winners. For scholarships, this could help find students who just miss the cut and suggest other schemes for them.This dual-methodology is highly relevant for a sophisticated scholarship system. While the Decision Tree provides a clear and interpretable prediction of a student’s eligibility, the FP-growth component offers a deeper understanding of the relationships between various attributes (e.g.,

IRJAEH
International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4274-4283
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0625

academic scores, extracurricular activities, financial background) that are common among award recipients. Kumari and Gopinath compared different machine learning models and found that Random Forest (a group of Decision Trees) worked best for both accuracy and reliability, even when the data was messy or incomplete [4]. This could be useful for scholarships where student data isn't always perfect.This is very important for scholarship systems, as application information can often have missing details. The study shows that using a model like Random Forest can help a system make good decisions even when the data isn't ideal. Yagcı showed that accurate predictions can be made using˘ only academic scores and institutional details, without using personal details like gender, caste, or religion.[5] This is important for Kerala's scholarships, as it protects privacy and reduces the chance of bias. By focusing only on relevant educational data, the model avoids unfair decisions that could happen if sensitive information is misused. The study also proves that removing personal details does not lower the accuracy of predictions, meaning fairness can be achieved without sacrificing performance. For scholarship recommendation systems, this method ensures that decisions are based purely on merit and academic performance, helping to build trust among students. Overall, the findings support the use of decision trees and similar models to give recommendations that are fair, clear, and reliable.

3. Addressing Fairness, Privacy, and Data Challenges

For a scholarship system to be truly fair, it must do more than just be accurate. It should treat all students equally, protect their privacy, and use data in an ethical way. This is especially important in Kerala's merit and minority scholarships, where differences in background, culture, and religion can create both obvious and hidden biases. Choosing the right data, avoiding unfair rules, and keeping personal information safe are key to building trust with students. Karthikeyan looked at how welfare schemes [6], including scholarships, were given to fishing communities in Tamil Nadu. It found problems like poor record-keeping, lack of awareness, and slow administration. These issues are similar to challenges in Kerala, where some deserving students might miss out. A scholarship system should include proper document checks and awareness programs so eligible students are not left out.The study highlights the need for better systems that keep accurate records, check documents carefully, and actively spread awareness about available scholarships. By fixing these issues, a scholarship system can make sure that help reaches the right people on time. Le Quy et al. tested different ways to measure fairness, such as making sure all groups get the same chance or that errors are equally spread.[7]They found that one model could look fair under one method but unfair under another. This means fairness is not a single fixed concept but depends on how it is measured. For scholarship systems in Kerala, this is an important lesson — the system should clearly state which definition of fairness it is using so that students and administrators understand the basis of the decisions. Being transparent about fairness rules helps avoid confusion and builds trust among applicants. Wakeel et al. [8] discussed how using prediction tools in education can help make policies more efficient but can also accidentally increase inequality. For scholarships, this means the system should include checks for bias, explain its decisions clearly, and give students a way to appeal if they think something is wrong.It should also explain its decisions in a way that students can easily understand, so they know why they were selected or rejected. Importantly, the system should provide a clear appeal process, allowing students to challenge a decision if they believe there was an error. These steps can help balance efficiency with fairness, making sure that technology supports equality rather than harming it. Andrewson et al. showed that carefully choosing and processing the right data makes models both more accurate and easier to understand. [9] For scholarships, this means the system can explain exactly how things like income, marks, and category affected the result. Their study showed that removing unnecessary or low-quality information reduces errors and helps the model focus on the factors that truly matter. For scholarships, this means the system can clearly explain how details like a student's family income, exam marks, and category affected the final decision. This makes the process transparent and helps students understand exactly why they were accepted or rejected. It also builds confidence in the system, since applicants can see that the decision was based only on relevant, reliable information rather than unrelated or biased factors. Febro found that using fewer but more important pieces of information makes predictions easier to explain and cheaper to run [10]. The study showed that when unnecessary details are removed, the system becomes simpler, faster, and more transparent, without losing much accuracy. In the context of scholarships, this means application forms can be shorter and still provide enough essential data—such as marks, income, and eligibility category—to make fair and well-informed decisions. A shorter form also reduces the burden on applicants and lowers the chance of errors or missing details. This approach can make the scholarship process more efficient and user-friendly while keeping the decision-making process clear and fair. Yagcı suggested using only academic and official school˘ data instead of personal details like religion or caste. [11] This protects privacy and reduces bias, which is important for Kerala's scholarships. By removing sensitive information, the risk of bias or discrimination is greatly reduced. This is especially important for scholarship systems in Kerala, where many schemes are aimed at promoting equality and fairness. Using only relevant academic and institutional data ensures that decisions are based purely on merit and objective criteria. It also builds trust among applicants, as they know personal characteristics unrelated to their ability or need will not influence the outcome. Chawla built a scholarship prediction model that looks at many factors — marks, family income, and activities outside class — and adjusts their importance based on each student's strengths. [12] The model is designed to adjust the importance of each factor depending on the student's individual strengths, allowing for a more balanced evaluation. This means that students who excel in areas other than academics—such as sports, arts, or community service—can still be recognized and rewarded. For Kerala, where scholarships are offered for a wide range of achievements and backgrounds, this approach is highly suitable. It ensures that opportunities are not limited to students with top grades alone, but also include those who demonstrate talent, leadership, or dedication in other fields. By using a flexible weighting system, the model supports fairer and more inclusive scholarship decisions.

4. Enhancing Performance with Advanced Techniques

Zeng and Acuna built a system called GOTFUNDING to recommend research grants by matching an applicant's profile with the best opportunities.[13]The system ranked the results so that the most relevant options appeared first, helping applicants focus on the best matches quickly. A key feature was its ability to explain why each match was made, such as how the applicant's experience, research area, or qualifications fit the grant requirements. This transparency allowed users to understand the reasoning behind the recommendations and increased trust in the system. A similar approach could be adapted for scholarships in Kerala, where students could receive a ranked list of the most suitable schemes along with clear explanations of why they qualify. This would make the process easier to navigate and ensure students do not miss opportunities that match their unique strengths and backgrounds. Sweeney et al. combined two models — Factorization Machines and Random Forests — to predict next-term grades.[14]This hybrid approach proved effective even when there was very little past information about a student, a challenge known as the "cold start" problem. By using the strengths of both models, the system could still make reliable predictions without needing a large history of academic records. For scholarships, this approach is especially useful when evaluating first-time applicants, such as new college students, who may not yet have an extensive performance record. It ensures that these students are judged fairly based on the information available, rather than being disadvantaged because of limited data. This method can help create a more inclusive scholarship system that gives equal opportunity to both returning and first-time applicants. Jayasree and Selvakumari created a hybrid neural network model that was more accurate than Decision Trees [15]. Even though neural networks are usually hard to explain, they used weight-visualization tools to make the results partly understandable. This allowed them to highlight which input factors—such as grades, attendance, or activities—had the most influence on the outcome. In a scholarship system, this kind of approach could be applied behind the scenes to ensure high accuracy in predictions. At the same time, a simpler and more transparent model, such as a Decision Tree, could be shown to the user to explain the decision in an easy-to-follow way. This combination would allow the system to deliver both strong performance and clear, trustworthy explanations to students. Khasne and Arjariya [16] used Correlation-based Feature Selection (CFS) together with an RNN-LSTM model to predict how students' learning outcomes change over time. CFS helped the system remove unnecessary or less useful information, keeping only the most important factors for accurate predictions. The RNN-LSTM model, which is designed to work well with time-based data, was able to identify patterns and trends in student performance across multiple terms or years. For scholarships that require students to maintain certain academic standards to continue receiving funding, this approach could be especially useful. It could help detect early warning signs for students who are at risk of losing eligibility, allowing interventions before the situation becomes serious. This would not only improve fairness but also give students a better chance to keep their scholarships. Zhidkikh et al. [17] found that looking at students' learning behaviors can predict dropout risk earlier than grades can. If scholarships want to keep students in school, adding these behavioral signals could help identify those who need extra support before they fail. For scholarship programs that aim not only to reward merit but also to support students in completing their education, these signals can be extremely valuable. By including behavioral data in the eligibility monitoring process, scholarships could identify students who might be struggling and offer extra help before they fail or drop out. This approach would make the system more proactive, ensuring that financial support is combined with timely academic or mentoring assistance to improve student success rates. Sekeroglu et al. studies on using machine learning to predict student success found that there's no standard way to measure performance, making it hard to compare models.Different studies used different evaluation metrics, such as accuracy, precision, recall, or F1-score, which makes it difficult to compare results fairly.[18] This lack of standardization means that a model that appears effective in one study might not perform as well when judged by another method. For scholarship systems, this finding highlights the importance of using clear, consistent, and widely accepted evaluation measures. Doing so would make the system's performance easier to explain, compare, and improve over time. It would also make the process more transparent and defensible if applicants or policymakers question how decisions are made. Kumar et al. [19] developed an AI-based scholarship finder that collects all scholarship information in one place and uses models like Naive Bayes and recommendation systems to match students with opportunities. One of its notable features is a data consistency checker, which verifies that the information a student provides such as income details, academic marks, and personal information is accurate and complete. This helps prevent mistakes or mismatches that could unfairly make a student ineligible. By combining matching algorithms with error checking tools, the system improves both the accuracy and fairness of scholarship recommendations. For a state like Kerala, this type of system could streamline the application process, reduce rejection rates caused by simple errors, and ensure that deserving students do not miss out on funding.

5. Decision-Making and Model Accuracy

Studies consistently show that interpretable models like decision trees achieve competitive accuracy while offering transparency. In cases where black-box models are used, post- hoc interpretability tools can bridge the gap. The ultimate choice of model should balance performance, fairness, and the ease of understanding for non-technical stakeholders. Wizsa and Rahmi [20] studied different methods for selecting scholarship recipients for an Indonesian government program. They compared several prediction models, including decision trees, boosting techniques, and logistic regression, to see which one performed best. Their results showed that logistic regression provided the highest accuracy while still being relatively simple to implement. The model could effectively identify which applicants met the eligibility criteria, making the selection process both

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4274-4283
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0625

fair and efficient. For scholarship systems in Kerala, a similar approach could be used to create a reliable and transparent prediction process. Logistic regression also has the advantage of showing how each factor—such as marks, income, or extracurricular achievements—affects the decision, which can make the process more understandable for applicants and administrators alike. Febri and Sari applied the Na¨ıve Bayes classifier to automate the selection process for Bank Indonesia scholarship recipients. The model achieved a high accuracy rate of 86.84%, showing that it could reliably identify eligible students [21]. Decisions were made based on clear factors such as GPA, parental income, and participation in organizations, making the process both transparent and data-driven. By using a statistical approach like Na¨ıve Bayes, the system could quickly evaluate large numbers of applications while ensuring fairness and consistency. For scholarship programs in Kerala, a similar model could help speed up selection, reduce human bias, and provide clear explanations of why each student was chosen or rejected. This would make the process more efficient while maintaining trust among applicants. Arcinas et al. developed a course recommendation system that applied several machine learning techniques to suggest the most suitable courses for students. Among the models tested, AdaBoost stood out as one of the most effective, achieving a remarkably high accuracy of 99.5%. [22] This demonstrated the model's ability to make highly reliable, data-driven decisions in an academic setting. While the study focused on course recommendations, the same approach could be applied to scholarship selection. By training an AdaBoost model on relevant eligibility criteria, the system could accurately match students to the most appropriate scholarships. Such a method would not only improve precision but also help ensure that decisions are consistent and based entirely on data, reducing the influence of bias or guesswork in the process. Ramaswami and Bhaskaran [23] showed that selecting a small set of highly relevant features improves predictive accuracy while reducing computational cost. For scholarship systems, focusing on key attributes such as academic performance, income level, and eligibility category ensures efficient, interpretable, and transparent recommendations [24-25].

3. Methodology

3.1 Data Collection

The proposed system will utilize the Kaggle dataset *Scholars.it*, which contains 245,671 records with attributes such as educational qualification, gender, community, religion, exservicemen status, disability, sports participation, annual percentage, family income, and eligibility outcome. This combination of historical data and official rules ensures that the system can learn from student patterns while complying with government requirements, shown in Figure 1.

1	Name	Education	Gender	Communit	Religion	Exservice-	Disability	Sports	Annual-Pe	Income	India	Outcome
2	INSPIRE Sc	Undergrad	Male	General	Hindu	Yes	Yes	Yes	90-100	Upto 1.5L	In	1
3	INSPIRE Sc	Undergrad	Male	General	Hindu	Yes	Yes	No	90-100	Upto 1.5L	In	1
4	INSPIRE Sc	Undergrad	Male	General	Muslim	Yes	Yes	Yes	90-100	Upto 1.5L	In	1
5	INSPIRE Sc	Undergrad	Male	General	Muslim	Yes	Yes	No	90-100	Upto 1.5L	In	1
6	INSPIRE Sc	Undergrad	Male	General	Chirstian	Yes	Yes	Yes	90-100	Upto 1.5L	In	1
7	INSPIRE Sc	Undergrad	Male	General	Chirstian	Yes	Yes	No	90-100	Upto 1.5L	In	1
8	INSPIRE Sc	Undergrad	Male	General	Others	Yes	Yes	Yes	90-100	Upto 1.5L	In	1
9	INSPIRE Sc	Undergrad	Male	General	Others	Yes	Yes	No	90-100	Upto 1.5L	In	1
10	INSPIRE Sc	Undergrad	Male	General	Hindu	Yes	Yes	Yes	90-100	Upto 1.5L	Out	0
11	INSPIRE Sc	Undergrad	Male	General	Hindu	Yes	Yes	No	90-100	Upto 1.5L	Out	0
12	INSPIRE Sc	Undergrad	Male	General	Muslim	Yes	Yes	Yes	90-100	Upto 1.5L	Out	0
13	INSPIRE Sc	Undergrad	Male	General	Muslim	Yes	Yes	No	90-100	Upto 1.5L	Out	0
14	INSPIRE Sc	Undergrad	Male	General	Chirstian	Yes	Yes	Yes	90-100	Upto 1.5L	Out	0
15	INSPIRE Sc	Undergrad	Male	General	Chirstian	Yes	Yes	No	90-100	Upto 1.5L	Out	0

Figure 1 Screenshots Showing Sample Dataset

3.2 Data Preprocessing and Cleaning

The dataset will be standardized to ensure consistency across categorical and numerical attributes. Missing values will be handled through imputation, categorical variables such as gender, community, and religion will be encoded numerically, and ranges (e.g., "90–100%" marks, "Upto 1.5L" income) will be converted into numeric bins. This step ensures the dataset is suitable for training machine learning algorithms.

3.3 Model Training and Eligibility Prediction

For the predictive component of the system, the XGBoost (Extreme Gradient Boosting) algorithm is selected due to its strong performance on structured, tabular data and its ability to handle noisy, incomplete, and heterogeneous inputs. Unlike a single decision tree, which may suffer from overfitting, XGBoost employs an ensemble of boosted trees that build sequentially, correcting errors from previous iterations. This approach consistently achieves high accuracy in educational data mining tasks while maintaining computational efficiency. The target variable is the eligibility outcome (eligible = 1, not eligible = 0), and the dataset is divided into training (70%), validation (15%), and testing (15%) subsets. Model performance will be assessed using standard metrics such as accuracy, precision, recall, and F1-score. The trained model serves as the eligibility predictor, forming the decision-making core of the scholarship recommendation system.

3.4 Rule Verification

Predictions generated by the model will be cross-verified against the rules table to ensure compliance with official criteria. For example, even if the model predicts eligibility for the State Merit Scholarship, the system will confirm that the student has at least 50% marks and family income below *Rs.*1,00,000. This hybrid approach ensures fairness and compliance with government guidelines.

3.5 Explainability

The proposed system emphasizes transparency in decisionmaking. Although gradient boosting models like XGBoost are complex ensembles, their predictions can be made interpretable through SHAP (SHapley Additive exPlanations). SHAP assigns each feature a contribution value for an individual prediction, showing how inputs such as academic percentage, family income, or category influenced the eligibility decision. These attributions are converted into natural language feedback (e.g., "Not eligible because family income exceeded Rs. 1,00,000 and academic percentage was below 50%"). By combining XGBoost with SHAP, the system offers both high predictive performance and transparent, studentfriendly explanations, ensuring fairness and trust in scholarship allocation.

3.6 Recommendation Engine

If a student is found ineligible for one scholarship, the system will not stop at rejection. Instead, it will match the student's profile with other available scholarships and generate a ranked list of alternatives based on eligibility conditions such as marks, income, and category. This ensures students are guided towards alternative opportunities rather than receiving a binary rejection.

3.7 System Flow

The overall workflow of the proposed system is shown in Figure 2. Student data is collected and preprocessed, followed by eligibility prediction through classifiers. Predictions are then verified against the rules table, and human-readable explanations are generated. When required, the recommendation engine suggests alternative scholarships, providing actionable and transparent outcomes, shown in Figure 2.

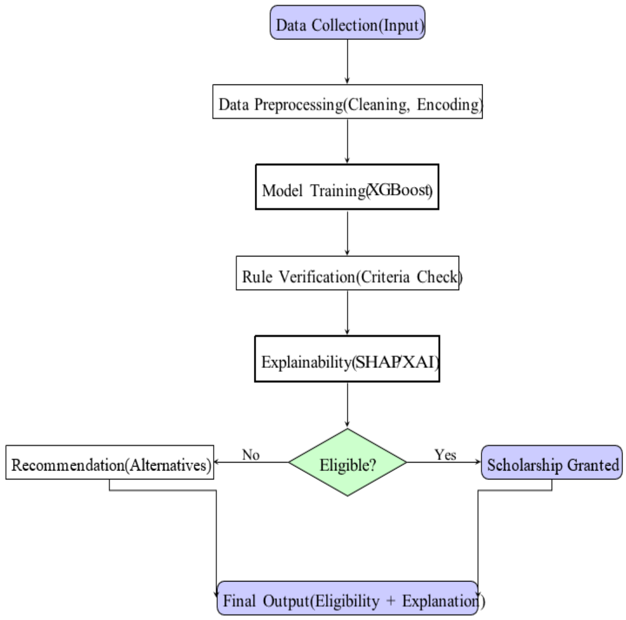

Fig. 2. Flowchart of the proposed scholarship eligibility prediction system.

Figure 2 **Flowchart**

4. Model Architecture and Working of XGBOOST

The architecture and working of the proposed XGBoost based scholarship eligibility prediction model are illustrated in Figure 3. The system integrates machine learning with explainability and rule verification to ensure fairness, transparency, and compliance with official government criteria.

Input and Preprocessing: The model receives cleaned data consisting of academic and socio-economic features such as marks, income, community, religion, gender, disability, and sports participation. Categorical features are numerically encoded, continuous attributes are normalized, and missing values are imputed.

Principle of XGBoost: XGBoost (Extreme Gradient Boosting) is an ensemble technique based on sequential learning of weak decision trees. Each new tree learns from the residual errors of the ensemble built so far. At iteration t, the model adds a tree ht(x) that minimizes:

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4274-4283
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0625

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + h_t(x_i)) + \Omega(h_t),$$

where l is the differentiable loss and $\Omega(h_t)$ regularizes tree complexity.

Using a second-order Taylor expansion, XGBoost efficiently optimizes:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} [g_i h_t(x_i) + \tfrac{1}{2} h_i h_t^2(x_i)] + \Omega(h_t),$$

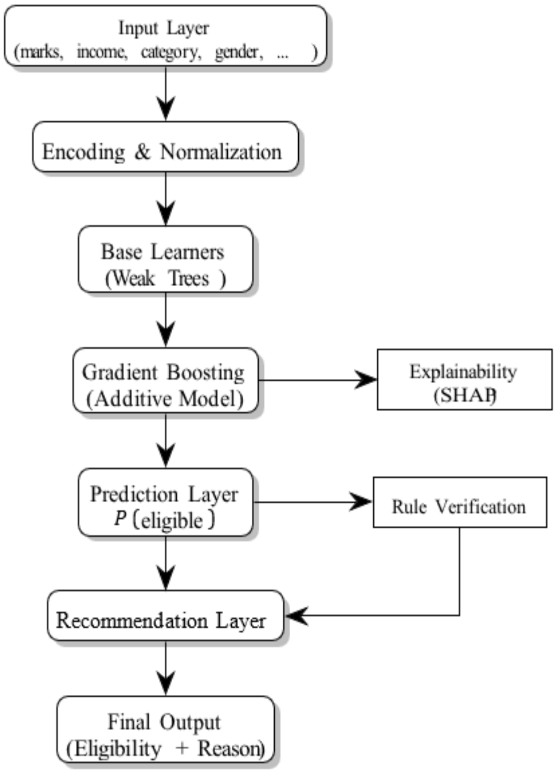

Figure 3 XGBoost-Based Scholarship Eligibility Model

Figure 3. Compact architecture of the XGBoost-based scholarship eligibility model. It combines boosted decision trees, explainability, and rule verification for transparent decisions.

where g_i and h_i are the first and second derivatives of the loss. This gradient-based optimization enhances both accuracy and speed.

Regularization and Optimization: XGBoost applies L1/L2 regularization, shrinkage (learning rate η), and subsampling to prevent overfitting and improve generalization.

Prediction Layer: The final eligibility probability is computed as:

$$\hat{y}_i = \sigma\left(\sum_{t=1}^{T} \eta\, h_t(x_i)\right),$$

where σ maps the score to [0,1] and a threshold (0.5) yield *Eligible/Not Eligible*.

Explainability and Rule Verification: The SHAP module assigns each input feature a contribution value, generating human-readable explanations. The rule verification layer crosschecks the model's decision with government-defined eligibility criteria.

Recommendation Layer: If ineligible, the system suggests alternative scholarships based on similarity matching. The proposed idea represents a novel step toward transforming scholarship recommendation systems into transparent, data-driven, and user-centric tools. By integrating explainable AI with rule verification, the model not only enhances accuracy but also simplifies implementation for educational institutions. This framework has strong potential for largescale deployment across states, enabling easy customization for diverse eligibility rules and criteria. In the long term, this invention can evolve into a unified, intelligent scholarship management platform that promotes fairness, inclusivity, and accessibility in education, marking a significant advancement in the automation and transparency of public welfare systems.

Conclusion

This study highlights the limitations of existing scholarship recommendation systems, particularly their reliance on binary acceptance or rejection outcomes without providing explanations. In the context of Kerala, where multiple merit and minority-based schemes involve complex and overlapping rules, students often face challenges in understanding eligibility criteria and identifying suitable opportunities. To address these issues, the proposed methodology combines machine learning classifiers with a rule verification layer and explainability tools. By integrating decision paths and SHAP values, the system not only predicts scholarship eligibility but also communicates the reasoning behind decisions in a transparent and user-friendly manner. The approach emphasizes fairness, interpretability, and compliance with official rules while also supporting personalization. Future work will focus on implementing the framework in Python with tools such as Scikitlearn, Pandas, and Streamlit, and evaluating its performance on Kerala-specific datasets. With these enhancements, the system has the potential to become a unified, explainable, and studentcentric platform for scholarship allocation.

References

[1]. H. Roslan and C. Chen, "A review of educational data mining methods for predicting student performance," *Journal of Computing*, vol. 15, no. 3, pp. 20–35, 2022.DOI:10.3991/ijet.v17i05.27685

[2]. S. Khan *et al.*, "Decision tree-based student performance monitoring system," *International Journal of Modern Education and Computer Science*, vol. 12, no. 8, pp. 55–64, 2021.

[3]. V. Bhegade and M. Shinde, "Combining decision trees and FP-growth for student scholarship prediction," in *Proc. International Conference on Computer Science and Information Technology*, pp. 112–117, 2016.

[4]. P. Kumari and S. Gopinath, "Random forest for scholarship recommendation with incomplete data," *Journal of Artificial Intelligence and Applications*, vol. 9, no. 4, pp. 112–120, 2022.

[5]. E. Yagcı, "Privacy-preserving prediction of student success using aca-˘ demic and institutional data," *Journal of Information and Privacy*, vol. 10, no. 1, pp. 45–56, 2022.DOI:10.1186/s40561-022-00192-z

[6]. R. Karthikeyan, "Welfare schemes and record-keeping in fishing communities of Tamil Nadu," *Journal of Social and Economic Development*, vol. 32, no. 2, pp. 89–101, 2018.

[7]. H. Le Quy *et al.*, "Measuring fairness in machine learning models for educational applications," *IEEE Transactions on Learning Technologies*, vol. 15, no. 4, pp. 501–510, 2022.

[8]. F. Wakeel *et al.*, "Bias and inequality in predictive tools for education policy," *Journal of Educational Technology and Society*, vol. 26, no. 1, pp. 123–135, 2023.

[9]. T. Andrewson *et al.*, "Data preprocessing for explainable predictive models," *Journal of Data Science and Analytics*, vol. 8, no. 3, pp. 150–162, 2021.

[10]. D. Febro, "The impact of feature reduction on explainability and cost of predictive models," *International Journal of Computer Engineering and Technology*, vol. 11, no. 5, pp. 88–97, 2020.

[11]. E. Yagcı, "Privacy-preserving prediction of student success using aca-˘ demic and institutional data," *Journal of Information and Privacy*, vol. 10, no. 1, pp. 45–56, 2022.

[12]. V. Chawla, "An adaptive AI model for multi-factor scholarship prediction," *IEEE Access*, vol. 12, pp. 101–110, 2024.

[13]. L. Zeng and M. Acuna, "GOTFUNDING: A recommendation system for research grants with explainable outcomes," in *Proc. ACM Conference on Recommender Systems*, pp. 45–54, 2022.

[14]. P. Sweeney *et al.*, "Hybrid models for predicting next-term grades with sparse data," *Journal of Educational Data Mining*, vol. 8, no. 2, pp. 78–90, 2016.

[15]. K. Jayasree and V. Selvakumari, "Hybrid neural networks for student performance prediction with explainability," *Neural Computing and Applications*, vol. 35, no. 12, pp. 8001–8015, 2023.DOI:10.3390/engproc2023059025

[16]. V. Khasne and D. Arjariya, "Predicting learning outcomes with CFS and RNN-LSTM," *Journal of Intelligent Systems*, vol. 2, no. 1, pp. 112–125, 2023.

[17]. V. Zhidkikh *et al.*, "Using behavioral signals to predict student dropout risk," in *Proc. International Conference on Educational Technology*, pp. 210–218, 2021.

[18]. B. Sekeroglu *et al.*, "A review of performance metrics for student success prediction models," *Journal of Educational Technology and Research*, vol. 17, no. 4, pp. 321–335, 2021.DOI:10.3390/app112210907

[19]. V. Kumar *et al.*, "An AI-based scholarship finder with data consistency checks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 2, pp. 212–225, 2024.

[20]. U. A. Wizsa and A. Rahmi, “Classification of KIP-K scholarship using logistic regression, classification trees, and boosting based on decision support system,” *Mathline: Jurnal Matematika dan Pendidikan Matematika*, vol. 10, no. 1, p. 837, 2025.DOI: https://doi.org/10.31943/mathline.v10i1.837
[21]. A. Febri and D. P. Sari, “Determination of Bank Indonesia scholarship recipients using Na¨ıve Bayes classifier,” *Barekeng: Jurnal Ilmu Matematika dan Terapan*, vol. 7, no. 2, pp. 11–20, 2023.DOI: https://doi.org/10.30598/barekengvol17iss3pp1595-1604
[22]. M. E. Arcinas *et al.*, “Predicting student’s academic performance using multi-layer perceptron,” *AIP Conference Proceedings*, vol. 3179, no. 1, p. 070008, 2025.DOI: https://doi.org/10.1063/5.0210008
[23]. M. Ramaswami and R. Bhaskaran, “A study on feature selection techniques in educational data mining,” *Journal of Computing*, vol. 1, no. 1, pp. 7–16, 2009. DOI: https://doi.org/10.48550/arXiv.0912.3924