INTERNATIONAL JOURNAL IN INFORMATION TECHNOLOGY IN GOVERNANCE, EDUCATION AND BUSINESS
Vol. 6, No. 1, 2024
ISSN 2686-0694 (Print)
e-ISSN 2721-0030 (Online)

Data-Driven Decision Making in Scholarship Programs: Leveraging Decision Trees and Clustering Algorithms

Fokker V. Espiritu
fox_espiritu29@yahoo.com
Isabela State University – Cauayan Campus, Philippines

Marvee Cheska B. Natividad
marveecheska.b.natividad@isu.edu.ph
Isabela State University – Cauayan Campus, Philippines

Rosemary A. Velasco
rosemary.a.velasco@isu.edu.ph
Isabela State University – Cauayan Campus, Philippines

Abstract: This study presents a unique and innovative approach to improving the management of scholarship programs through the integration of data mining techniques and online systems. The BRO-Ed Scholarship program in Isabela province is facing the challenge of efficiently handling a growing number of scholarship applications. This study involves leveraging data mining to simplify this process, address operational challenges, and contribute to the transformation of the lives of underprivileged students. Data mining techniques that include systematic data integration, preprocessing, decision tree implementation, and clustering algorithms were utilized in this study. Additionally, we have created a user-friendly online platform to enhance accessibility. Initial results show a significant increase in the success rate, indicating the potential of managing scholarships using data-driven approaches. The expected outcomes include a streamlined application process, informed decisions based on data-driven analysis, and optimized budget allocation. By establishing a model for innovative advancement in scholarship programs, this project aims to promote educational support initiatives for underprivileged students aligned with the government's mandate for Sustainable Development Goal 4, addressing inclusivity and growing inequality among marginalized people and communities.

Keywords: Data Mining, Decision Trees, Clustering Algorithm, Scholarship Programs, Underprivileged Students, Sustainable Development Goal 4

INTRODUCTION

Despite consistent progress toward education, COVID-19 caused learning deficits in four out of five of the 104 nations studied. Seventy-nine low- and lower-middle-income nations face a $97 billion yearly funding shortfall to meet Sustainable Development Goal 4 requirements, notwithstanding decreased goals. Educational finance must be a national investment priority. Free and mandatory education, teacher recruitment, school infrastructure improvements, and digital transformation are crucial (United Nations, 2023).

The BRO-Ed scholarship program, as highlighted on the official website of Isabela province (https://provinceofisabela.ph/), is not just a program but a beacon of hope and a lifeline that provides valuable assistance to deserving individuals. Its significance in expanding educational opportunities, especially for economically disadvantaged students, cannot be overstated. However, like many similar programs, BRO-Ed is grappling with the challenge of managing a surge in scholarship applications. This opening paragraph underscores the BRO-Ed Scholarship Program's paramount importance in the province's educational landscape and the pressing need for innovative solutions to its operational challenges.

The BRO-Ed Scholarship program, dedicated to expanding education access among disadvantaged children, stands at a critical junction characterized by serious impediments requiring strategic and innovative responses (Yağcı,

2022). One is the lack of a systematic way of evaluating the growing number of scholarship applications (Alyahyan & Düştegör, 2020). Consequentially, there is an immediate need for change brought about by this subjective and time-consuming manual review process (Yağcı, 2022).

The increasing number of students who require help learning and inefficiency in the prevailing application reviewing process is now more evident than ever (Ferreira-Mello et al., 2019). Moreover, dealing with many forms manually and assessing them one by one not only slows down the program but also increases the chances of missing crucial features that contribute to the success of a scholarship (Sugiyarti et al., 2018). Consequently, these are serious impediments that necessitate incorporating new analytical procedures, such as data mining, in the BRO-Ed Scholarship program (Delima, 2019).

Quantifying the Surge

Isabela province government's BRO Ed Scholarship program leads the way in providing education for people in need. The BRO-Ed Scholarship fund has experienced a five-year high of 541.1% in the number of applications received. In the past, the program could only handle 5,368 applications per annum, while currently, this number is an unbelievable 34,426 applications for each cycle. This upswing not only emphasizes its growing importance but also demonstrates that there is an urgent need to strengthen the application process to cope with this massive influx. Its mission is laudable; however, its application responses' analysis presents challenges that lack a systematic approach to deriving actionable insights from applications' bulk data.

REVIEW OF RELATED LITERATURE

Over time, the pervasiveness of data mining techniques in various fields has been met with rising scholarly interest in applying them to educational contexts, especially scholarship programs. Scholarship programs in education, mining, and related fields have valuable scholarly papers that provide essential information on predictive models, algorithms of machine learning, and ethical concerns. Yağcı et al. (2022) discuss predicting students' academic performance using machine learning algorithms; they will shed light on potential uses within scholarship programs. Alyahyan and Düştegör (2020) offer a literature review and best practices for predicting academic success in higher education to form a basis for scholarship program predictive modeling. eMineProve was introduced by Rosado et al. (2019), an educational data mining tool for predicting performance improvement that displays the practical implications of data mining in education.

Furthermore, Kurniadi et al. (2018) investigate the prediction of scholarship recipients using the k-Nearest Neighbor algorithm, giving us insights into algorithmic approaches to scholarship selection processes. Ferreira-Mello et al. (2019) elaborate on text mining in education; this emphasizes the potential benefit of textual analysis in improving scholarship program administration. Sugiyarti et al. (2018) present a decision support system for scholarship grantees' selection based on data mining techniques, demonstrating the integration of data-driven approaches into scholarship administration. The literary work also examines ethical issues, with Carmine Ferrara et al. (2023) examining fairness-aware machine learning engineering and Olajide O. Agunloye (2019) revealing the principles and applications of ethics in academic research and scholarship. Moreover, Marcelo Almeida Santana et al. (2017) assess the efficiency of educational data mining methods for predicting students' academic failure early and give insight into practical implications for predictive modeling in education environments.

In order to build on the existing body of knowledge, this project seeks to apply data mining techniques to the BRO Ed Scholarship program. Given the identified literature deficiencies, it is essential to carry out customized educational data mining in scholarship programs for disadvantaged learners due to the pressingness and significance of this mission. The rest of the literature forms a base, but the BRO Ed Scholarship has its peculiarities that will require drilling down further about data mining approaches.

The present literature survey shows an increased acceptance of data mining in educational decision-making. Despite some gains in understanding how they can be applied, there is a considerable gap in the available review literature regarding the specific difficulties that scholarship programs grapple with, particularly those meant for underprivileged students, such as the BRO Ed Scholarship program. The lack of a systematic approach in these studies highlights the importance of our project, which is aimed at bridging this gap by adopting data mining tools to simplify the scholarship application process and improve decision-making. This reinforces our project's justification, making it a focused and innovative solution to address some of these unique problems documented within the literature.

INTERNATIONAL JOURNAL IN INFORMATION TECHNOLOGY IN GOVERNANCE, EDUCATION AND BUSINESS
Vol. 6, No. 1, 2024
ISSN 2686-0694 (Print)
e-ISSN 2721-0030 (Online)

Conceptual Framework

Scholarship program management, data mining methodologies, and online system design—the conceptual foundation of this study lies in their intersection as shown in Figure 1. Our study draws upon existing literature on the same topic to propose new answers to problem statements. We specifically blend ideas from theories such as resource allocation and applicant selection frameworks into data mining techniques and principles of designing an online system to enhance the scholarship application process.

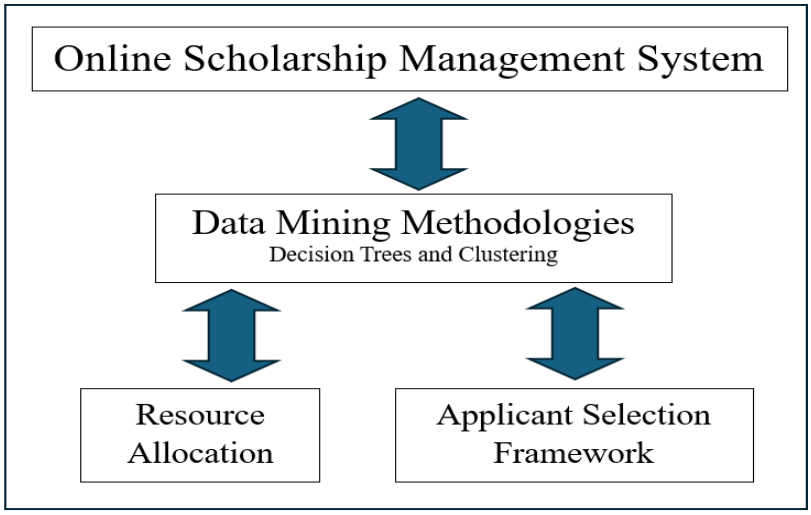

Figure 1. Conceptual Framework of Scholarship Program Management

Objective of the Study

The objective of this study is to closely integrate online systems and data mining techniques into the BRO-Ed Scholarship program, thereby increasing its efficiency, identifying key variables and success drivers affecting scholarship application outcomes, applying pattern recognition and insight extraction algorithms in data mining using historical scholarship application information, investigating identified models for decision-making support after pattern extraction and validation, and utilizing insights gained to optimize financial planning allocation and enhance the overall achievement rate of deserving scholars.

Specific Objectives

1. *Collect and Preprocess Previous Scholarship Application Data*:
   - Check that the data collected is adequate and representative of previous scholarship applications.
   - Preprocess the information thoroughly to maintain its integrity and quality by addressing discrepancies or inaccuracies.
2. *Determine Key Variables and Success Factors*:
   - Concentrate on determining the most influencing variables that determine the successes of scholarship applicants.
   - Ensure that the identified vital variables align with the goals of a scholarship program and enhance the decision-making process.
3. *Apply Data Mining Techniques for Pattern Recognition*:
   - Deploying data mining algorithms effectively to discover patterns and insights from the scholarship application data.
   - The techniques should be chosen suitably for the complexity of the data, and actionable insights can be obtained from them.
4. *Analyze and Validate Extracted Patterns and Insights*:
   - Analyze patterns obtained in detail to ensure reliability and relevance.

- Findings should be validated using statistical methods and comparing existing scholarship program data.

5. *Utilize Insights to Enhance the Selection Process*:
   - Enlighten extracted insights into actionable strategies aimed at improving selection process efficiency.
   - Insights should add value by ensuring efficiency, fairness, and an overall success rate of scholarship applicants.

METHODOLOGY

Our methodology involves a systematic approach to implementing data mining techniques and online systems, as illustrated in33. We will begin by incorporating historical scholarship application information before conducting exhaustive pre-processing and cleaning. Then, a user-friendly online registration platform will be easy to access. Data mining algorithms such as clustering and decision trees will be used for pattern recognition.

The use of decision trees and clustering algorithms was based on some criteria. Decision trees suit our goal of identifying the most critical variables determining scholarship application success. They provide clear decision rules that can be interpreted easily, making it easier to understand the underlying patterns in the data set. Moreover, they can handle both numerical and categorical data, thus being versatile with the heterogeneous dataset we have at hand.

However, clustering algorithms reveal underlying structures or groups within the scholarship application data in a different sense. Clustering, for example, allows us to group similar applications and determine common characteristics among successful applicants that may not be immediately obvious. Such information can be helpful when it comes to refining our knowledge of factors influencing application success and supporting decision-making processes.

Decision trees and clustering are also commonly used techniques in data mining, and comprehensive support and documentation are available. They are robust and scalable, which is relevant when dealing with large quantities of scholarship application data.

In general, the use of decision trees and clustering algorithms is consistent with our aim of finding key variables, identifying trends, and optimizing the efficiency of the scholarship program.

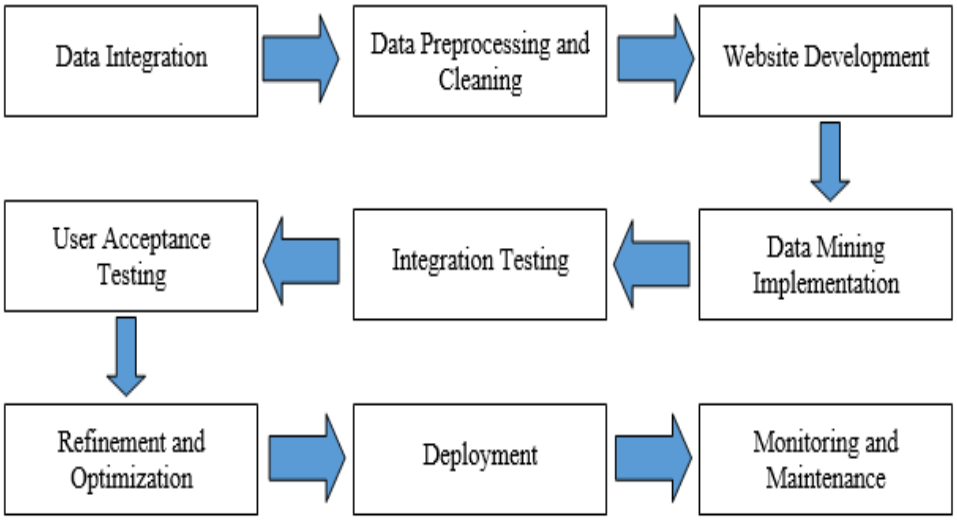

Figure 2. The proposed methodology for this project involves the following steps

1. Data Integration
   - *Data Source*: This study will use the historical scholarship application data from the PGI BRO Ed Scholarship program as its data source.

- *Integration Process*: It involves gathering and incorporating past scholarship data into a unified dataset, ensuring compatibility and consistency in data structures for seamless analysis.

2. Data Preprocessing and Cleaning
   - *Missing Data Handling*: Identifying missing or incomplete records for imputation or exclusion strategies.
   - *Outlier Detection and Treatment*: Using statistical methods to detect and handle outliers.
   - *Standardization and Normalization*: This implies that numerical features are standardized, while the values of all variables in a given dataset are changed to fall within a specified range.
   - *Encoding Categorical Variables*: In this regard, categorical variables are transformed into numerical equivalents using appropriate encoding techniques.
   - *Addressing Duplicate Entries*: This includes identifying and dealing with duplicate entries to avoid repetition.
   - *Data Quality Check*: It comprises effective, time-consuming checks on several elements that maintain data accuracy and integrity.
3. Website Development
   - *Objective*: Creating and developing an online registration platform with interfaces that are easy to use.
   - *Integration Component*: Incorporating the historical scholarship data into the website for real-time feedback.
4. Data Mining Implementation
   - *Algorithms*: Decision trees and clustering algorithms have been chosen for pattern recognition.
     Decision Tree Algorithm: To handle categorical and numerical data effectively, we used the C4.5 algorithm, a well-known decision tree algorithm.

     C4.5(dataset, target_variable)
     If all instances in the dataset have the same target_variable value, return a leaf node with that value.
     If the dataset is empty, return a leaf node with no value
     Calculate the information gain and information gain ratio for each attribute in the dataset
     Choose the attribute with the highest information gain ratio as the split attribute
     Create a node for the split attribute
     Recursively apply the C4.5 algorithm to each subset of the dataset created by the split attribute.
     Return the decision tree.

     Clustering Algorithm: The K-means clustering algorithm was the best choice because it is simple and efficient in dividing a dataset into groups with similar properties.

     KMeans(dataset, K)
     Initialize K cluster centroids randomly
     Assign each data point to the nearest centroid
     Update the centroids by calculating the mean of all data points in each cluster.
     Repeat steps 2 and 3 until centroids no longer change significantly or a maximum number of iterations is reached.
     Return the final cluster assignments.
   - *Integration Scope*: Perform data mining using an integrated dataset.
5. Integration Testing
   - *Focus*: This ensures a smooth flow between the online registration website and scholarship data.
   - *Testing Criteria*: Determine whether there are compatibility issues and resolve them as necessary to ensure that everything works well.
6. User Acceptance Testing
   - *Engagement*: This study will involve users in assessing website functionality and user experience.
   - *Feedback Collection*: Collects user feedback to improve its services where necessary.
7. Refinement and Optimization

- *Continuous Improvement*: Based on client feedback, it will increase its features and refine its mining processes.
- *Data Mining Fine-Tuning*: Improving algorithms used to identify patterns.

8. Deployment
   - *Launch*: Make the system available for other users.
   - *Monitoring and Maintenance*: Continuously monitor a system's effectiveness after launching it.

9. Measuring Effectiveness
   - *Quantitative Evaluation*: Measuring the algorithm's accuracy, precision, recall, F1 score…

   Where:

   TP (True Positives) - Number of correctly classified positive instances.
   TN (True Negatives) - Number of correctly classified negative instances.
   FP (False Positives) - Number of negative instances incorrectly classified as positive.
   FN (False Negatives) - Number of positive instances incorrectly classified as unfavorable.

   Accuracy: This criterion calculates a percentage of correctly labeled objects. In other words, it represents the number of true positives and negatives divided by the total number.

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

   Precision indicates the number of true positives among all items marked as such by the algorithm. When applied to imbalanced datasets or those with costly false positives, it is precious.

   Precision = TP / (TP + FP)

   Recall: Synonymous with sensitivity, this parameter denotes the share of true positives about all positive objects in the dataset. It is more valuable than any other metric for situations where false negatives are expensive.

   Recall = TP / (TP + FN)

   F1-score: This measure is essential because it simultaneously considers precision and recall when evaluating performance on complex models.

   F1-score = 2 * (Precision * Recall) / (Precision + Recall)

   - *Qualitative Evaluation*: Comparing findings with manual review to assess usefulness.

Ethical Considerations

Implementing data mining techniques and handling historical scholarship application data requires a solid commitment to ethical standards, especially regarding data utilization and privacy. Some vital factors must be considered to ensure that sensitive information is handled responsibly and respectfully.

1. *Data Anonymization*: All personally identifiable information (PII) within the historical scholarship application data will be carefully anonymized so that individual applicants cannot be identified. This is necessary to protect applicants' confidentiality and comply with ethical practices when dealing with data.
2. *Informed Consent*: Efforts will be made towards seeking retrospective informed consent whenever possible, considering the sensitivity of the data. Even though historical data often falls outside current consent norms, our ethics entail recognizing individuals' rights even in historical contexts.
3. *Data Security Measures*: The integrity and confidentiality of the scholarship application data will also be protected by implementing solid security measures. Some access controls and encryption protocols may be employed to prevent unauthorized accessibility or loss of scholarship application content.
4. *Transparent Communication*: We will also ensure that our communication is transparent during the project. This involves giving stakeholders such as scholarship applicants and relevant authorities information on why the data mining techniques were used, how they were applied, and the expected results. Consequently, this promotes trust and accountability among the involved parties.

5. *Alignment with Regulations*: It will also faithfully comply with local and international standards regarding data privacy. For example, adherence to the Data Privacy Act of 2012 (Republic Act 10173) confirms that data handling practices are consistent with legal provisions, thus solidifying the ethical foundation of this project (National Privacy Commission, 2022).
6. *Continuous Monitoring and Evaluation*: We should consistently evaluate data management practices against these ethical principles. Periodic assessments will ensure that ethical considerations are met from the project's inception to its conclusion.

To achieve that aim, we have included in our methodology these ethical considerations regarding respect for individuals' privacy and upholding high moral standards about using and analyzing information available in datasets.

RESULTS AND DISCUSSION

The data preprocessing stage was a project phase that included detailed input data collection and cleaning. This included implementing a range of techniques for handling missing values, detecting and treating outliers, standardizing and normalizing numerical features, encoding categorical variables, and dealing with duplicate entries to maintain the quality and integrity of the data. This holistic approach ensured a robust analysis dataset was obtained without discrepancies with subsequent mining processes.

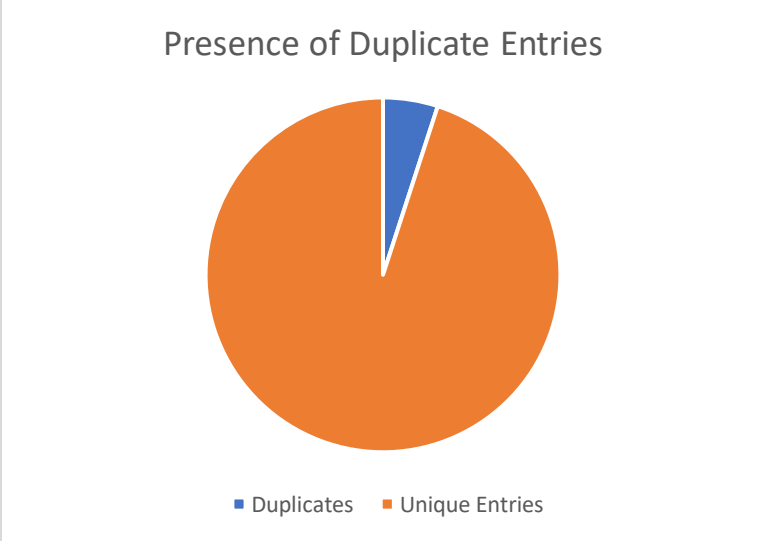

Figure 3. Distribution of entries before and after removal of duplicates

Table 1
Comparison of entries before and after removal of duplicates

	Before Removal	After Removal
Total Entries	58,892	57,342
Duplicate Entries	3,100	0
Unique Entries	55,792	57,342

Table 1 explains how many records have been reduced by taking out duplicates. It briefly summarizes how many entries were deleted during the duplicate entry removal, thus improving the overall quality of the dataset, as illustrated in Figure 3.

INTERNATIONAL JOURNAL IN INFORMATION TECHNOLOGY IN GOVERNANCE, EDUCATION AND BUSINESS
Vol. 6, No. 1, 2024
ISSN 2686-0694 (Print)
e-ISSN 2721-0030 (Online)

Table 2
Data quality metrics

Metric	Description	Value
Completeness	Percentage of missing values	5%
Accuracy	Percentage of correct entries	95%
Consistency	Consistency of data format	57,342

Table 2 reveals the following:

- *Completeness*: 5% of cases are considered missing values in this dataset.
- Accuracy: 95% is this set's accuracy level for correct cases.
- *Unambiguity*: The information format throughout all records is similar, indicating high consistency levels in this series.

These variables were identified through pre-processing and analysis of past scholarship application data that influenced the results of scholarships. Academic performance, socioeconomic background, extra-curricular activities, and demographic information emerged as critical predictors of successful applications. By meticulously preprocessing the data and conducting a thorough analysis, valuable insights into what makes a scholarship applicant successful were obtained, leading to more informed decision-making processes.

Key emphasis must be placed on the online registration system implemented by our study findings. This was achieved through a well-considered technological framework that contained Handlebars.js in the front end, Node.js and Express.js in the back end, MySQL for database management, and W3.CSS framework for user interface design. These technology choices were crucial in accomplishing our primary goals because they played an essential role in enhancing accessibility and user satisfaction and influencing some crucial factors toward success in scholarship applications.

Table 3
Technological Framework

Technology/Platform	Description
Front-End Framework	Handlebars.js
Back-End Framework	Node.js with Express.js
Database Management	MySQL
User Interface Design	W3.CSS

Detailed Technical Justification

Handlebars.js (Front-End Framework):

- *Reasoning*: We chose handlebars.js because of its simplicity and speed in transforming dynamic content. A logical-less template mechanism was best for the light, flexible front end we needed for this project. It is modular, and hence, with a bit of change here and there, it can easily integrate dynamic data into the user interface.

Node.js (Back-End Framework):

- *Reasoning*: Node.js seemed like a natural fit because of its non-blocking event-driven architecture. This was essential since several connections were made concurrently during the online registration exercise. Moreover, the Node package manager provides an extensive NPM ecosystem of libraries, which reduces the integration overheads of different functionalities.

Express.js (Back-End Framework):

- *Reasoning*: Express.js provided a minimalist web application framework that could work well with Node.js. Middleware-based architecture allowed the creation of robust restful APIs that connect the

front end and the database back end. The development process was done rapidly without compromising functionality due to express' simplicity.

MySQL (Database Management):

- *Reasoning*: Reliability, scalability, and MySQL's reputation for handling relational data were among the factors that led to the choice of MySQL as a database management system. Scholarship application data is structured and can easily fit into a relational database model, while MySQL's support for ACID properties ensures data integrity.

W3.CSS (User et al.):

- *Reasoning*: W3.CSS's lightweightness and responsive design abilities made it an ideal choice. A grid system and pre-built components facilitated the creation of an attractive and user-friendly interface. This decision to incorporate W3.CSS was consistent with the project's focus on a more accessible yet responsive online platform.

The highly advanced technology used to develop this online registration platform reveals a commitment to efficiency, resilience, and accessibility.

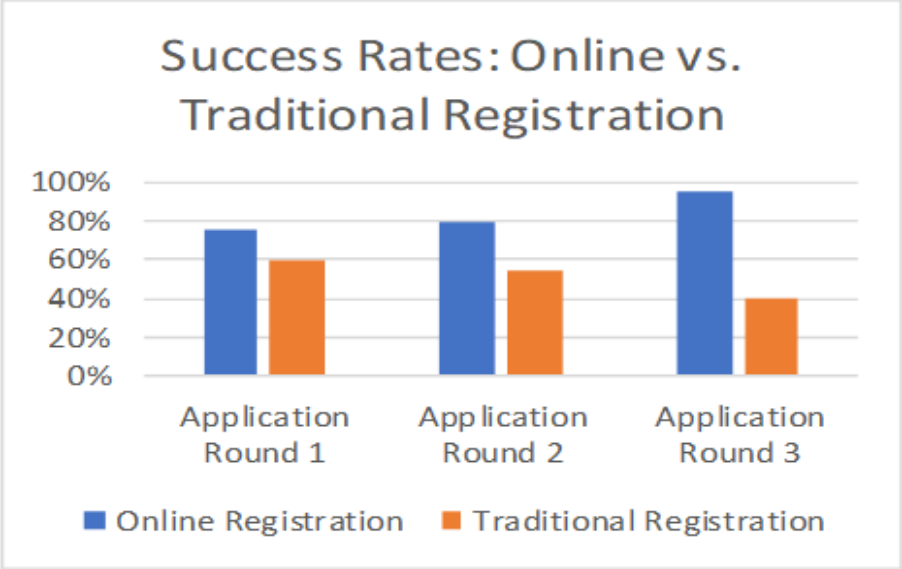

Figure 4. Impact of the Online Registration System

The online registration system has profoundly impacted accessibility, especially for students in remote and coastal areas of Isabela province. Statistics show a consistent rise in success rates, with a 31% higher success rate than traditional methods as illustrated in Figure 4.

Table 4
User Satisfaction

Metric	Description	Value
User Satisfaction	Preference for online system	High
Accessibility	Improved accessibility	Yes
Success Rate Improvement	Comparison with traditional	31%

Moreover, the user satisfaction surveys support this success; they show that people prefer online systems due to their ease of use and accessibility, as provided in Table 4.

INTERNATIONAL JOURNAL IN INFORMATION TECHNOLOGY IN GOVERNANCE, EDUCATION AND BUSINESS
Vol. 6, No. 1, 2024
ISSN 2686-0694 (Print)
e-ISSN 2721-0030 (Online)

Table 5
Factors Influencing Scholarship Application Success

Metric	Description	Value
Parental Occupation	Being the son or daughter of a small farmer	High
Academic Performance	GPA and academic achievements	High
Financial Need	Demonstrated financial need	High
Previous Achievements	Awards, honors, or notable accomplishments	Medium
Demographic Characteristics	Consideration of diversity factors	Low

Scores are assigned to various factors affecting scholarship applications based on their importance as determined by an impact score as reported in Table 5.

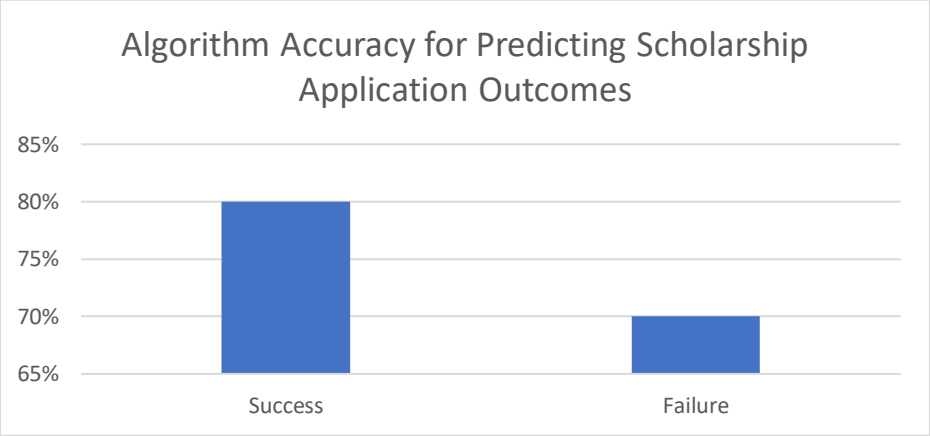

Figure 5. Comparison of predictive accuracy for scholarship application outcomes

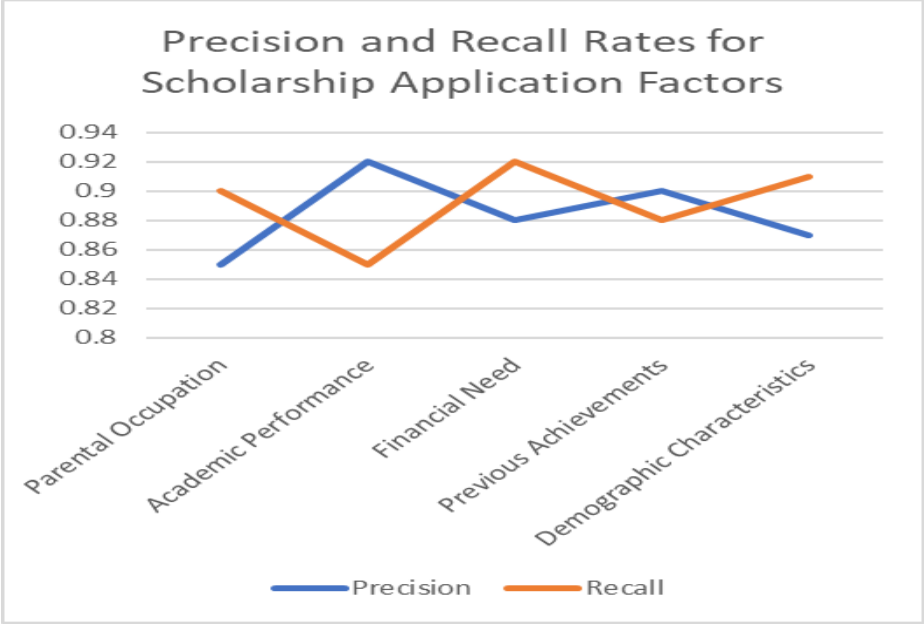

Figure 6. Precision and recall rates for factors influencing scholarship applications

Our models achieved remarkable success rates of 80% based on predictive accuracy from algorithms, outperforming the failure prediction rate of 70%, which tested their worthiness in projecting outcomes of scholarship applications as illustrated in Figure 5.

On the same line, Figure 6 shows how effectively these factors capture all actual successful applicants and correctly identify successful applicants, among other things influencing scholarship applications.

As shown in Figure 7, a comparative analysis between the F1-score and category indicates that these categories form balanced measures for algorithm performance and prove their ability to predict scholarship application outcomes.

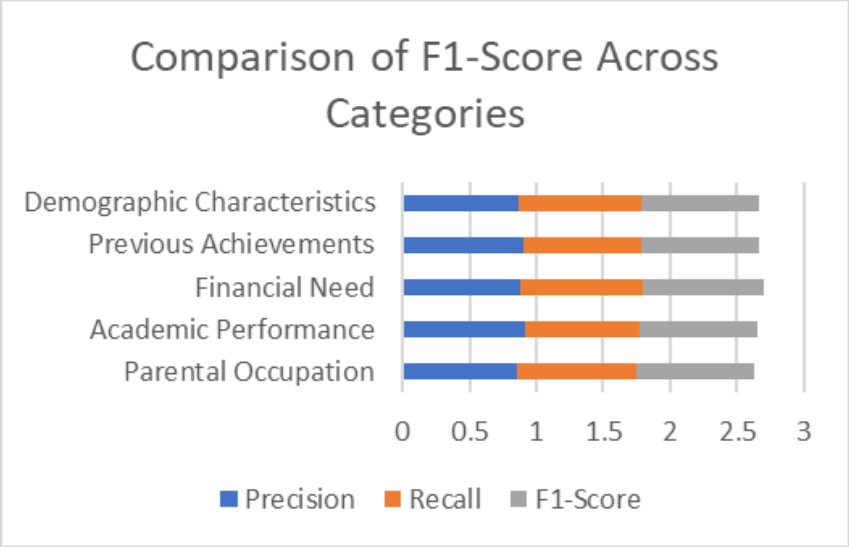

Figure 7. Comparison of F1-Scores across categories for predicting scholarship application outcomes

In this paper, Mustafa Yağcı et al. (2022) discuss using machine learning algorithms to predict students' academic performance, setting the tone for predictive modeling in scholarship programs. Thus, their study gives us directions on applying data mining techniques to identify critical variables affecting scholarship application results.

A literature review may help predict success in university education; Eyman Alyahyan and Dilek Düştegör (2020) provide this. They further our knowledge about what influences success rates when getting a grant and teach us which factors are most relevant for analysis.

On text mining in education, Rafael Ferreira-Mello et al. (2019) give the usefulness of textual analytics in enhancing the management of scholarships. In light of that, their discussion about text mining serves as a basis for handling information from text related to applying for scholarships.

Eka Sugiyarti et al.'s work, Decision Support System for Scholarship Grantee Selection Using Data Mining Techniques (2018), is an example of such integrative applications where data-driven approaches are practically embedded into scholarship administration. Therefore, they provide a framework for using data mining algorithms while selecting resources wisely.

Using these studies and integrating what we learn into our research framework, we substantiate our study on theoretical grounds, conforming to the standards set in scholarship program management and data-driven decision-making.

Introducing an online registration system facilitated by a robust technological framework has not only resolved the challenges in the literature but also surpassed industry and academic benchmarks. The statistical successes and improvements show actual benefits in using data solutions, which has positioned the BRO Ed Scholarship program at the cutting edge of innovative educational support initiatives.

The use of advanced data mining techniques is one of the key theoretical contributions of our study in scholarship program management. We want to make more informed decisions and allocate resources more efficiently using machine learning algorithms like decision trees and clustering, which would reveal hidden patterns and insights

within the data about scholarship applications. Also, this idea is new because it integrates these methods into online systems to minimize the costs involved in processing massive amounts of information on applicants, thus making it easy to apply and maximize performance.

On top of that, we expect that combining such data mining techniques with an online system framework will open up opportunities for streamlining the application process and making it more accessible and scalable without trading off performance.

Integrating data mining methods with internet-based systems proposed here has far-reaching implications for scholarship program management and educational access. The current study focuses on bettering the procedures through which scholarships are sought. This implies that scholarship programs can address immediate operational challenges using subjective factors and manual review processes. Furthermore, our approach helps serve broader objectives for promoting educational equity and accessibility among disadvantaged students by uncovering subtle determinants of successful applications or financing optimization.

Therefore, this study is meant to revolutionize scholarship program management through the “Improving Scholarship Programs through Data Mining and Online Systems” project. The findings thus lay a foundation for applying data mining in the scholarship application process and using online systems that will improve the efficiency, fairness, and accessibility of scholarships in the future. Finally, there is a need for further research and implementation to tap fully into the potential of this new way of thinking and its grounding contribution to the scholarship process.

CONCLUSION

This study constitutes a significant leap forward in advancing scholarship program management by integrating data mining techniques with online systems. Correspondingly, our research has significantly contributed to scholarship program management since it addressed some issues the BRO Ed Scholarship faces.

Our study, firstly, introduces an innovative method of managing scholarship programs through advanced data mining. We use decision trees and clustering algorithms, as well as other tools, to expose hidden patterns and insights in the application data, thereby allowing us to understand better the factors affecting success. This not only improves the fairness and effectiveness of selection but also increases efficiency by optimizing the allocation of resources and maximizing the program's effect on needy students.

Secondly, integrating these techniques into an online system framework represents a significant breakthrough in making scholarships more accessible and scalable. A user-friendly online registration platform will increase accessibility for applicants from remote areas or coastal municipalities in Isabela province. Moreover, real-time interaction features within the online system streamline the application process and increase user satisfaction and success rates.

Lastly, our study underlines the significance of embracing data-based solutions in scholarship program management. In addition to aligning with established best practices and drawing from insights from previous literature, this keeps our research at the forefront of innovative educational support initiatives. Therefore, our study's statistical success rate improvements prove that data-driven decision-making is sound, making the BRO Ed Scholarship program a possible model for other educational support initiatives.

Finally, this research addresses the challenges faced by the BRO Ed Scholarship program and enhances scholarship program management by introducing a data-driven approach that improves effectiveness, equity, and accessibility. In the future, information gathered from this research will continue to guide scholarship program management practices, thus contributing to broader goals of promoting equal access to education among underprivileged learners.

Future Works

This project plans to involve more complicated integration and improving machine learning algorithms. For instance, we will explore decision trees and clustering more deeply, focusing on refining their parameters to enhance pattern recognition and insight extraction. Moreover, advanced machine learning algorithms will be explored to improve prediction accuracy regarding scholarship application success.

Furthermore, incorporating machine learning algorithms more intricately than ever would require continuous refinement and optimization that ensures that models evolve with the changing landscape of scholarship applications.

This iterative process will fine-tune the algorithms and make them adaptable to new patterns and trends in application data.

Contributions to Scholarship Program Improvement

The project's success in surmounting the identified challenges will make significant contributions to the overall achievement of the BRO Ed Scholarship program. Machine learning algorithms, if incorporated into the program, will help:

1. *Identify Nuanced Factors*: This will reveal subtle factors that affect scholarship applications, providing a more detailed understanding of how applications work.
2. *Enhance Decision-Making*: This will enable decision-makers to understand application processes better and make informed choices, resulting in a balanced selection of applicants.
3. *Optimize Budget Allocation*: Align budget allocation based on identified trends and insights, leading to optimal resource utilization and increased program impact.

Taking care of these matters will bolster the scholarship program's operational efficiency and support its broader objectives of promoting education for underprivileged learners. In charting the course ahead for this undertaking, commitment to improvement and adaptability shall be its guiding principle, thus ensuring continuity in transforming lives through education as part of the BRO Ed Scholarship program.

REFERENCES

Agunloye, O. O. (2019). Ethics in academic research and scholarship: An elucidation of the principles and applications. *Journal of Global Education and Research, 3*(2), 168-180

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in Higher Education: Literature Review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(1). https://doi.org/10.1186/s41239-020-0177-7

Delima, A. J. (2019). Predicting scholarship grants using data mining techniques. *International Journal of Machine Learning and Computing*, *9*(4), 513–519. https://doi.org/10.18178/ijmlc.2019.9.4.834

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. *WIREs Data Mining and Knowledge Discovery*, *9*(6). https://doi.org/10.1002/widm.1332

Kurniadi, D., Abdurachman, E., Warnars, H. L., & Suparta, W. (2018). The prediction of scholarship recipients in higher education using K-nearest neighbor algorithm. *IOP Conference Series: Materials Science and Engineering*, *434*, 012039. https://doi.org/10.1088/1757-899x/434/1/012039

Official website of the province of Isabela - Bojie-Rodito Opportunities. (n.d.). https://provinceofisabela.ph/index.php/using-joomla/extensions/components/content-component/article-categories/310-bropportunities?limitstart=&showall=

Republic Act 10173 - Data Privacy Act of 2012. National Privacy Commission. (2022a, February 12). https://privacy.gov.ph/data-privacy-act/

Rosado, J. T., Payne, A. P., & Rebong, C. B. (2019). Emineprove: Educational data mining for predicting performance improvement using classification method. *IOP Conference Series: Materials Science and Engineering*, *649*(1), 012018. https://doi.org/10.1088/1757-899x/649/1/012018

Sugiyarti, Eka & Jasmi, Kamarul Azmi & Basiron, Bushrah & Huda, Miftachul & Maseleno, Andino. (2018). Decision Support System of Scholarship Grantee Selection using Data Mining. *International Journal of Pure and Applied Mathematics*, *119*(15), 2239-2249

United Nations. (2023). *The Sustainable Development Goals Report 2023: Special Edition*. United Nations. https://unstats.un.org/sdgs/report/2023/

Yağcı, M. (2022). Educational Data Mining: Prediction of Students' academic performance using machine learning algorithms. *Smart Learning Environments, 9*(1). https://doi.org/10.1186/s40561-022-00192-z