ASHESI

Title: Machine Learning Model that Supports Scholarship Decision-Making at Ashesi University

Course: Intro to AI

Cohort C

Date: April 30, 2025

Lecturer: Owusu Asamoah Dennis

Group 5 Members:

1. Matthew Tuurozeeng
2. Benjamin Mutie Charles
3. Alan Kofi Safo Ofori
4. Hassan Maltiti Yakubu

1. Organisation and Problem Background

Ashesi University is one of Africa's renowned institutions emphasising educational access through its Financial Aid Office. Each year, the office receives hundreds of scholarship applications from talented students with diverse backgrounds and financial situations. Determining who is most deserving of aid involves sifting through essays, academic records, recommendation letters, family income statements, leadership experience, and extracurricular activities, which are time-consuming and subjective.

Given Ashesi's mission to promote equity and excellence through ethics, entrepreneurship and leadership, the Financial Aid Office faces a delicate balancing act: ensuring no promising student is left behind while maintaining transparency and fairness in allocation.

However, with growing application volumes and a capped required number and financial resources each year, there is a need for a decision-support tool that is scalable, consistent, and efficient.

The central challenge, then, is this: How can the Financial Aid Office streamline the scholarship evaluation process without compromising on fairness or quality of judgment?

2. How machine learning solves the problem

The project introduces a machine learning (ML) based scholarship eligibility prediction system, which is designed to act as a complementary decision-support tool. Instead of replacing the part of the human judgment, the model improves it by providing data-driven insights that identify students with a high likelihood of receiving financial aid for their undergraduate studies.

3. Problem Statement:

Design and develop a machine learning model to predict scholarship eligibility based on historical application data, helping the Financial Aid Office prioritise candidates more efficiently and fairly.

The model uses ensemble learning (a combination of multiple models: logistic regression and random forest) trained on applicant features such as GPA, family income, WASSCE scores, leadership scores, essay ratings, extracurricular participation, recommendation strength, and whether the student is a first-generation university attendee. The model uses probability-based outputs that predict eligibility, which the financial aid office can use to support final decisions.

The ML system was deployed through a user-friendly Streamlit web app that allows staff to input applicant data and receive instant eligibility feedback with confidence scores.

4. Challenges faced during data collection and preprocessing

One of the significant challenges was access to realistic data from the financial aid office. Because of data privacy issues, actual student scholarship application data could not be used. As a result, a synthetic dataset was created to mimic the structure and distribution of real-world data. However, this posed several difficulties:

- *Balancing the target variable (eligible):* The initial dataset was imbalanced with more "not eligible" cases, which made biased early model predictions. Instead of applying SMOTE (synthetic minority oversampling technique), it was addressed by generating a larger, more balanced synthetic dataset. This helped improve the model's ability to generalise and reduce bias when making predictions.
- *Feature selection:* Variables like essay scores and recommendation strength must be quantified meaningfully. Categorical values were encoded to numerical representations (0s and 1s), but human interpretability was ensured to make sure outputs from the model can be interpreted by humans
- *Data scaling and model selection:* Models usually perform poorly due to the wide range of feature values (e.g., income vs. GPA). This prompted the necessity for normalisation and ensemble methods for improved performance.

5. Ethical Considerations

Building a model that predicts eligibility for scholarships raises ethical issues. While machine learning can enhance consistency, it also risks introducing hidden biases if the data it is trained on reflects existing inequalities. In this project:

- *Data privacy was respected:* No real student data was used, and any future use of this system should adhere to Ashesi's data protection policies and data protection laws in Ghana
- *Bias mitigation:* Care was taken to prevent the model from over-relying on features like family income or region, which could unintentionally disadvantage underrepresented groups.
- *Transparency and accountability:* The model does not make final decisions; it serves as a complement. The Financial Aid Office retains human oversight to ensure decisions remain fair, contextual, and values-aligned.

6. Conclusion

This project demonstrates how machine learning can support scholarship decisions in a fair, scalable, and transparent way. While not a replacement for human judgment, the tool helps streamline complex evaluations, reduce manual workload, and offers a starting point for further refinement. Continuous feedback can make this system valuable to Ashesi's financial aid process.

References

Authority, U. S. (n.d.). *Ethical considerations in the use of Machine Learning for research and statistics*. UK Statistics Authority. https://uksa.statisticsauthority.gov.uk/publication/ethical-considerations-in-the-use-of-machine-learning-for-research-and-statistics/pages/7/

ITLawCo. (2025, January 28). *Ghana's Data Protection Act 2012 (Act 843) - ITLawCo*. https://itlawco.com/focus-areas/data-protection-and-privacy/ghanas-data-protection-act-2012-act-843/#:~:text=Ghana%E2%80%99s%20Data%20Protection%20Act%2C%202012%20%28Act%20843%29%20establishes,features%20of%20the%20Act%2C%20and%20required%20compliance%20actions.

Kanwal, B., Shoukat, R. S., Rehman, S. U., Kundi, M., AlSaedi, T., & Alahmadi, A. (2024). A new framework for scholarship predictor using a machine learning approach. *Intelligent Automation & Soft Computing*, *0*(0), 1–10. https://doi.org/10.32604/iasc.2024.054645

Upchurch, T. (2018, April 8). To work for society, data scientists need a hippocratic oath with teeth. *WIRED*. https://www.wired.com/story/data-ai-ethics-hippocratic-oath-cathy-o-neil-weapons-of-math-destruction/