Beyond Performance: Explaining and Ensuring Fairness in Student Academic Performance Prediction

Istiak Alam

Department of CSE

January 28, 2026

Presentation Outline

- Problem Motivation
- Dataset Description
- Related Work and Limitations
- Limitations of the Proposed Paper
- Targeted Work and Future Scope
- Conclusion

Why This Study Is Important

- Most student performance prediction models focus only on accuracy
- Fairness and bias are often ignored
- Black-box models lack transparency
- Unfair predictions can negatively impact students

Core Argument:
High predictive accuracy alone is insufficient in educational decision-making.

Dataset Used in This Study

- Dataset: **UCI Student Performance Dataset**
- Source: UCI Machine Learning Repository
- Courses:
  - Mathematics (student-mat.csv)
  - Portuguese (student-por.csv)
- Total records after preprocessing: 395 students
- Total attributes: 33

Demographic Attributes

Attribute	Description
sex	Gender of the student
age	Age of the student
address	Urban or rural residence
famsize	Family size
Pstatus	Parents living together or apart

Note: These attributes may introduce bias and are treated as sensitive features.

Socioeconomic Attributes

Attribute	Description
Medu	Mother's education level
Fedu	Father's education level
Mjob	Mother's occupation
Fjob	Father's occupation
guardian	Primary guardian of the student

These attributes strongly influence academic performance but raise fairness concerns.

Academic and Behavioral Attributes

- studytime – Weekly study hours
- failures – Number of past failures
- schoolsup – School educational support
- famsup – Family educational support
- paid – Extra paid classes
- absences – Total absences
- goout – Social activity level
- Dalc, Walc – Alcohol consumption
- health – Health condition

These features are actionable and useful for early interventions.

Target Variable

- G1 – First period grade
- G2 – Second period grade
- G3 – Final grade (Target)

Prediction Tasks:

- Classification: Pass / Fail
- Regression: Continuous grade prediction

Overview of Related Work

- Traditional ML models (LR, SVM, RF)
- Deep learning-based approaches
- Focus primarily on predictive accuracy
- Limited attention to fairness and explainability

Key Related Studies and Limitations

- Cortez & Silva (2008)
  - Used decision trees and regression
  - Limitation: No fairness or bias analysis
- Albreiki et al. (2021)
  - Compared multiple ML models
  - Limitation: Ignored demographic bias
- Livieris et al. (2021)
  - Semi-supervised learning
  - Limitation: Lack of interpretability

Research Gap Identified

Main Gap

No existing work simultaneously optimizes:

- Prediction accuracy
- Fairness across sensitive attributes
- Model explainability
- Reproducibility

Dataset Limitations

- Dataset collected only from Portugal
- Limited cultural and geographic diversity
- No ethnicity or income attributes
- Results may not generalize globally

Methodological Limitations

- Use of SMOTE-NC introduces synthetic data
- Binary pass/fail classification loses grade granularity
- Limited fairness metrics used
- No real-world deployment validation

How This Paper Addresses Past Limitations

- Introduces fairness-aware evaluation
- Uses explainable AI techniques (SHAP, LIME)
- Balances accuracy with ethical considerations
- Uses open datasets for reproducibility

Future Research Directions

- Multi-country educational datasets
- Intersectional fairness analysis
- Causal fairness modeling
- Teacher-in-the-loop evaluation
- Longitudinal performance tracking

Conclusion

- Accuracy alone is insufficient in educational AI
- Fairness and explainability are essential
- The paper provides a strong baseline
- Significant opportunities remain for future improvement

Thank You