Beyond Performance: Explaining and Ensuring Fairness in Student Academic Performance Prediction with Machine Learning

...

Paper: Beyond Performance: Explaining and Ensuring Fairness in Student Academic Performance Prediction with Machine Learning
github: https://github.com/kadirkesgin/fairsynedu

Core Idea of the Paper

The paper proposes a holistic machine learning framework for predicting student academic performance that balances prediction accuracy, fairness, and interpretability, ensuring that models do not disadvantage students based on gender or socioeconomic background .

Key Research Questions

- Can student academic performance be accurately predicted using open-access educational data?
- How do different ML models (Logistic Regression, Random Forest, XGBoost) compare in performance?
- Do these models introduce bias with respect to gender and socioeconomic attributes?
- Can explainable AI (SHAP, LIME) and fairness-aware techniques mitigate such biases without harming accuracy?

Dataset Used

- UCI Student Performance Dataset
- Merged from student-mat.csv and student-por.csv
- 395 unique students, 33 features

- Includes demographic, academic, behavioral, and socioeconomic attributes
- Target variable: Final grade (G3), binarized as pass/fail

Why This Matters

Educational ML systems can reinforce social inequalities if fairness is ignored. Schools increasingly rely on AI for early intervention and decision-making. The study shows how to build responsible, ethical AI for education rather than just high-accuracy models .

Preprocessing Pipeline

- Merge datasets and remove duplicates
- One-hot encode categorical features
- Standardize numerical features
- Handle class imbalance using SMOTE-NC
- Train–test split (80/20) with stratification
- 5-fold cross-validation for model training

Key Insight

Past academic performance (G1, G2) and absences dominate prediction outcomes. Model complexity does not guarantee better performance—simpler models can generalize just as well. Bias becomes much stronger at intersectional levels (e.g., gender + parental education) .

Key Result

Logistic Regression achieved performance comparable to XGBoost and Random Forest. Adversarial debiasing reduced:

- Demographic Parity gap: $0.048 \rightarrow 0.021$
- Equalized Odds gap: $0.298 \rightarrow 0.180$

Only a small drop in accuracy, showing fairness–performance trade-off is manageable .

Critical Insight

Fairness problems may appear minor for single attributes, but become severe when attributes intersect (e.g., gender + low parental education), which many prior studies fail to analyze .

Major Strengths

- Unified pipeline combining performance, fairness, and explainability
- Uses open-access data → reproducible research
- Strong use of SHAP + adversarial debiasing
- Practical relevance for real educational systems

Major Limitations

- Dataset limited to Portuguese secondary schools → low generalizability
- Heavy reliance on G1 and G2 grades, which are unavailable in early prediction scenarios
- SMOTE may introduce synthetic bias artifacts
- No deep learning or longitudinal modeling explored

Extension suggestions

1: Comparative Explainability

What others do:
Use SHAP or LIME

What YOU will do:
Use both SHAP and LIME

Compare:

- Feature importance consistency
- Stability across models
- Interpretability for educators

This becomes a research question, not just implementation.

Example research question:

Which explainability method provides more stable and actionable insights for student performance prediction?

2: Model–Explanation Alignment

What others do:

Train model

Explain model

Stop

What YOU will do:

Analyze whether explanations align across models

Logistic Regression vs Random Forest vs XGBoost

Identify features that consistently matter

This shows scientific thinking, not just coding.

3: Actionable Educational Insights

This is where many theses fail — you won't.

What others do:

"Study time is important" (obvious)

What YOU will do:

Map explanations to real interventions

Attendance → counseling

Assignment delay → mentoring

Low engagement → early warning

This turns the work into decision-support research.

4: Feature Reduction with Explainability

Your contribution:

Use SHAP to identify top-k features

Retrain models using only those features

Compare:

- Accuracy
- Interpretability

- Deployment simplicity

  This is strong, practical, and publishable.