Report Submission: Higher Education Student Performance Prediction System

1. Feature Engineering Decisions

During the development process, we deliberately avoided creating additional feature columns due to the following considerations: - New features introduced multicollinearity between existing features - The addition of new features led to model performance degradation even using synthetic smoted data - The original feature set provided sufficient predictive power without additional complexity - Feature interactions were already captured in the base features

2. Data Preprocessing Limitations

The preprocessing phase faced certain limitations: - Limited availability of external data for merging - Data collection was region-specific, making cross-region merging potentially problematic - Focus on maintaining data integrity and relevance to the specific use case - Privacy concerns limited access to additional student information

3. Oversampling Technique Changes

The project underwent a significant change in the oversampling approach: - Initially implemented ADASYN for handling class imbalance - Switched to SMOTE due to the following error: `"Error in preprocessing training data: Not any neighbours belong to the majority class. This case will induce a NaN case with a division by zero. ADASYN is not suited for this specific dataset. Use SMOTE instead."` - SMOTE provided more stable results and better handling of the minority class

4. Model Performance Observations

Important findings regarding synthetic data usage: - When using SMOTE for training and validation splits, the classification metrics showed promising results - However, performance degraded significantly when tested on real-world data - This observation aligns with the known limitations of synthetic data in machine learning, where models may perform well on synthetic test sets but fail to generalize to real-world scenarios - The gap between synthetic and real-world performance highlights the importance of proper validation strategies

---

Generated on: 27-03-2025