**Early Warning Academic Performance Prediction**

**Dual-Path Ensemble Framework with Integrated Analytics Interface**

**Executive Summary**

This project implements a novel approach to academic early warning systems through a dual-path ensemble model that combines traditional machine learning with temporal sequence analysis. The implementation uniquely integrates conversational AI capabilities with advanced visualization, creating an accessible yet powerful platform for educational decision-makers.

**Primary Objectives**

1. Develop a dual-path ensemble model that effectively predicts student academic outcomes using:
   1. Demographic/static features processed through random forest algorithms
   2. Temporal engagement patterns analyzed through sequential models (GRU)
2. Create an equitable early warning system that maintains consistent performance across demographic groups, with particular attention to socioeconomic indicators (IMD bands)
3. Implement a hybrid visualization framework that combines traditional dashboard elements with conversational analytics
4. Demonstrate technical integration between Python modeling pipelines and enterprise analytics platforms

**Novel Technical Architecture**

This project introduces several innovative elements to the educational analytics domain.

***Dual-Path Model Integration***

Unlike traditional approaches that treat demographic and behavioral data separately, this architecture processes these distinct data types through specialized pathways before merging.

A diagram of a model of a structure

AI-generated content may be incorrect.

**Demographic Path Processing:**

* One-hot encoding of categorical variables
* Feature importance analysis
* Handling missing demographic data

**Temporal Path Processing:**

* Time-window feature extraction
* Sequence normalization
* VLE engagement patterns

The complete system architecture, including ConverSight integration, is visualized in the accompanying diagram (see System Architecture Diagram section).

***Hybrid Analytics Interface***

The project implements a combination of traditional dashboards and conversational AI interface:

* Traditional visualizations: Fixed views showing key metrics and trends
* Conversational layer: Natural language queries allow non-technical users to explore predictions
* Cross-platform integration: Enables seamless transitions between visual and conversational interfaces

***Fairness-Aware Implementation***

The project incorporates equity considerations directly into the modeling pipeline:

* Stratified sampling across demographic groups
* Balanced performance metrics with emphasis on IMD bands
* Fairness-aware model selection criteria

**Concrete Deliverables**

***Python Implementation***

* Feature engineering pipelines for both model paths
* Random Forest implementation for demographic features
* GRU implementation for temporal engagement patterns
* Ensemble integration module with optimization
* Evaluation framework with equity metrics
* API endpoints for external integration

***Visualization System***

* Tableau dashboards showing:
  + Risk distribution across modules and demographics
  + Temporal engagement patterns
  + Assessment completion metrics
  + Model performance indicators
* Integration with ConverSight's Athena for:
  + Natural language querying of prediction
  + Automated anomaly detection
  + Contextual narrative generation

***Documentation***

* Technical architecture documentation
* Model validation reports
* Demographic fairness analysis
* Implementation guide
* API documentation

**Implementation Timeline (Feb 25 - Apr 11, 2025)**

|  |  |  |
| --- | --- | --- |
| **Dates** | **Phase** | **Activities** |
| Week 1 | Data Preparation | * Data exploration * Feature engineering pipeline * Data preprocessing |
| Week 2 | Model Development | * Random Forest path implementation * GRU path implementation * Basic ensemble integration |
| Week 3 | Model Optimization | * Hyperparameter tuning * Ensemble optimization * Fairness evaluation |
| Week 4 | Core Visualization | * Tableau dashboard development * Key metric implementation * Interactive filtering |
| Week 5 | ConverSight Integration | * Athena integration * Natural language interface * Basic automated insights |
| Week 6 | Validation & Documentation | * Cross-validation * Demographic fairness testing * Technical documentation * Final presentation preparation |

**Significance**

This project advances educational analytics in several ways:

* Methodological Innovation: The dual-path ensemble approach specifically addresses the mixed-data-type challenge in educational datasets by processing each through optimized pathways.
* Technical Architecture: The integration of traditional machine learning with conversational AI represents an emerging approach in educational technology, making advanced analytics accessible to non-technical users.
* Equity Focus: By embedding demographic fairness directly into the modeling process, this project addresses a gap in current early warning systems, which often show variable performance across demographic groups.
* Implementation Approach: The hybrid visualization strategy combines the strengths of established visualization tools with emerging conversational interfaces, creating a more accessible and comprehensive system.

**Technical Requirements**

***Development Environment***

* Python 3.8+ with scikit-learn, TensorFlow, pandas, numpy
* Git/GitHub for version control
* Tableau Desktop/Public for visualization
* ConverSight.ai account (limited integration)

***Data Requirements***

* OULAD dataset with all component files
* Preprocessing scripts for temporal feature extraction
* Assessment data normalization

***Deployment Considerations***

* Local development environment with Anaconda
* GitHub repository for code management
* Static exports for visualization demonstration

**System Architecture Diagram**

The system architecture diagram illustrates the integration between the Python modeling components, Tableau dashboards, and the ConverSight platform. Key elements include (1) data flow from source systems through the Python modeling pipeline, (2) dual processing paths for demographic and temporal data, (3) integration points between Tableau visualizations and ConverSight components, and (4) user interface options including both visual dashboards and natural language queries.

A diagram of a system

AI-generated content may be incorrect.

The diagram demonstrates how the system combines traditional analytics approaches with conversational AI to create a comprehensive early warning system.

**Conclusion**

This project implements a technically sophisticated yet practical approach to early warning systems in education. By combining advanced machine learning techniques with emerging conversational interfaces, it demonstrates both technical proficiency and an understanding of real-world educational challenges. The dual-path architecture specifically addresses the complex nature of educational data, while the hybrid interface makes these insights accessible to educational decision-makers.