Predictive Analytics and Multi-Agent Orchestrator for Educational Performance

Project Report Submitted in Partial Fulfilment of the Requirements for the Degree of

Bachelor of Technology in Computer Science and Engineering

By

Archi Mangla (2210110180)
Nehal Pruthi (2210110430)

Under the Supervision of

Dr. Sonia Khetarpaul
Associate Professor, Department of Computer Science and Engineering

Dr. Pooja Singh
Assistant Professor, Department of Computer Science and Engineering

SHIV NADAR
INSTITUTION OF EMINENCE DEEMED TO BE
UNIVERSITY
DELHI NCR

Shiv Nadar University, Delhi NCR
Department of Computer Science and Engineering

November, 2025

Abstract

This report presents an end-to-end educational intelligence platform built from a merged SAT, ACT and TSI dataset (2020–2023). We construct a robust data foundation (cleaning, encoding, derived 3-year trends and campus/district/region deviations), an Analytics Agent that computes a normalized Academic Risk Index and equity-gap diagnostics, and a Prediction Agent that imputes masked values and forecasts next-year readiness using ensemble regressors (Random Forest / XG-Boost) and optional sequence models. A What-If Agent performs policy simulations (e.g., raise participation or score targets) to quantify downstream effects on readiness and risk, and an Intervention Agent produces prioritized, evidence-based recommendations (rule-based + LLM-enhanced) to close gaps. We validate model performance with regression and classification metrics, visualize trends and gap heatmaps, and demonstrate scenario-driven sensitivity analysis. The platform is designed for operational use by district leaders to identify at-risk cohorts, evaluate intervention impact, and prioritize equitable strategies for improving college and career readiness.

Contents

1 Introduction and Motivation

Predicting student academic readiness has become a central priority for school districts striving to improve college and career outcomes. Standardized assessments such as the SAT, ACT, and TSI serve as key indicators of postsecondary preparedness, yet they also reveal persistent disparities across campuses, demographic groups, and regions. Administrators often face challenges in interpreting multi-year trends, quantifying risk, identifying equity gaps, and estimating the potential impact of policy decisions. Without a unified analytical framework, these insights remain fragmented and difficult to operationalize.

To address this need, we develop an integrated educational intelligence platform using combined SAT, ACT, and SAT/ACT–TSI readiness datasets from 2020–2023. By cleaning and merging three statewide assessment sources into a single longitudinal dataset, we create a reliable foundation for trend analysis and predictive modeling. Building on this, we design a multi-agent architecture that incorporates (i) an Analytics Engine to compute academic risk indices and benchmark comparisons, (ii) machine-learning models to impute masked values and forecast next-year performance, (iii) a What-If Engine that simulates policy scenarios such as increased participation or score improvements, and (iv) an Intervention Agent that recommends targeted, evidence-based strategies for campus leaders.

The motivation behind this work is twofold. First, districts require timely and interpretable data tools that can surface at-risk student groups before performance declines become entrenched. Second, decision-makers benefit from scenario-driven insights that demonstrate how specific improvements such as raising SAT Math scores or increasing participation would influence readiness, equity gaps, and future outcomes. By bringing together data engineering, predictive analytics, and LLM-driven recommendations, this platform aims to support more equitable, data-informed planning across schools and districts.

2 Related Work

Recent advances in Educational Data Mining (EDM) and learning analytics have demonstrated the value of machine learning for understanding and predicting student outcomes across large and complex datasets. Early foundational work established EDM as a discipline focused on extracting actionable patterns from educational data, emphasizing tasks such as prediction, clustering, and trend modeling [1]. Subsequent studies have shown that classical ML methods—including Decision Trees, k-NN, SVMs, Linear Regression, and Naïve Bayes—can effectively model academic performance, with ensemble and boosting methods often outperforming single-model approaches [2, 3]. More recent research highlights the increasing importance of transparency and interpretability, integrating Explainable AI (XAI) techniques such as SHAP and LIME to support stakeholder trust and decision-making [4]. Within this line of work, Al-Barrak and Al-Razgan (2016) applied decision-tree models to predict students' final GPA [5]; however, their contribution remained limited to predictive modeling without offering deeper analytics or actionable decision-support capabilities.

Parallel to prediction-focused EDM research, another stream of literature examines the broader role of machine learning in educational science, particularly as large-scale datasets emerge through digital assessments, MOOCs, and institutional data systems. These studies emphasize the shift toward scalable, data-driven evaluation pipelines and call for methodological frameworks that integrate ML into educational policy and institutional planning [6]. Complementing this, emerging research explores the role of Large Language Models (LLMs) and LLM-based Multi-Agent Systems (MAS) in educational contexts, where agents support pedagogical guidance, personalized feedback, and instructional decision-making [7, 8]. Although primarily oriented toward classroom-level interaction, this growing body of work underscores the potential of multi-agent, AI-driven systems for broader educational support.

Across these domains, prior work predominantly focuses on (1) course-level or student-level predictions, (2) demonstrating algorithmic performance on limited datasets, or (3) pedagogical applications of AI. In contrast, there is comparatively limited research on institution-scale analytics pipelines that integrate longitudinal datasets, forecasting, equity diagnostics, and policy simulation into a unified framework.

Our work extends this literature by combining three statewide datasets (SAT, ACT, TSI readiness) into a multi-year, campus-level panel and designing a comprehensive ML-augmented decision-support architecture. While prior studies highlight the value of ML predictions or XAI explanations individually, our system integrates (i) a composite Academic Risk Index, (ii) subgroup equity gap analysis, (iii) a What-If simulation engine, and (iv) an LLM-driven Intervention Agent into a cohesive four-agent platform. This shifts the use of ML in education from isolated predictive modeling to a systematic, actionable, institution-facing intelligence system designed to support campus- and district-level planning at scale.

3 Methodology

This project adopts a multi-stage methodological pipeline that transforms raw statewide SAT, ACT, and TSI data (2020–2023) into an integrated decision-support system for campuses and districts. The full workflow consists of five coordinated phases: data preparation, analytical computation, machine learning, policy simulation, and intervention generation. Together, these phases form a unified multi-agent architecture capable of delivering both retrospective analysis and forward-looking planning insights.

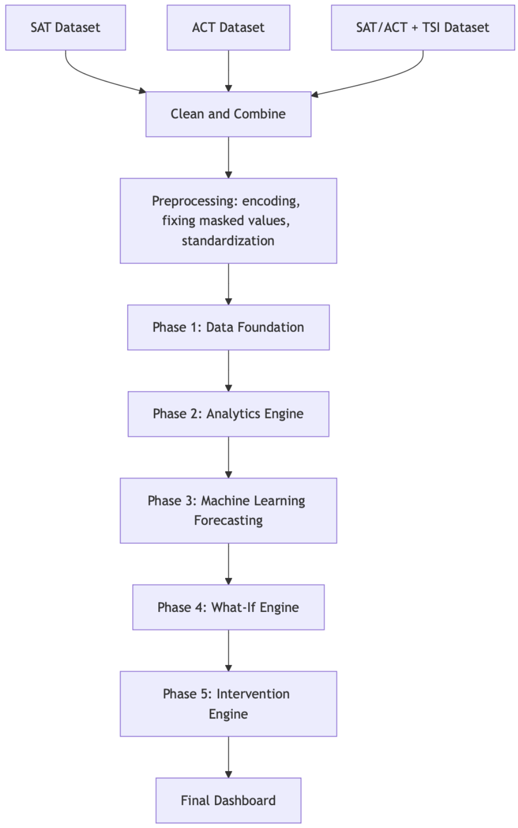

Figure 1: End-to-end workflow of the proposed AI-driven educational intelligence platform.

A) Data Foundation

1. Dataset Integration. Three independent statewide datasets (SAT, ACT, and SAT/ACT with TSI readiness) were standardized and merged into a single multi-year panel. Integration involved harmonizing campus, district, and regional identifiers; aligning variable names; and ensuring consistent year-level structure across all tables.

2. Data Cleaning and Encoding. Masked values (e.g., "¡25") were converted to numerical `NaN`, categorical fields such as *Group* were label-encoded, and score-related fields were normalized for consistency. Data quality checks resolved missing codes, duplicated records, and inconsistencies across years.

3. Constructing Derived Features. To enable robust analytics, the system generated derived indicators including rolling 3-year trends, year-over-year score changes, campus–district–region deviations, and group-level summary metrics. These engineered features form the basis of both the risk index and subsequent predictive models.

B) Analytics Engine

4. Academic Risk Index. A composite 0–100 risk score was constructed using indicators from all three assessment systems: SAT (ERW, Math), ACT (English, Math, Composite), and TSI readiness. Additional components include participation rates, multi-year performance decline, and subgroup equity gaps. All indicators were normalized using min–max scaling and integrated through a weighted scoring formula designed for interpretability and cross-campus comparability.

5. Equity Gap Analyzer. The system computes subgroup differences relative to campus-, district-, and region-level benchmarks across SAT, ACT, and TSI readiness metrics. It further captures 3-year subgroup trends to detect widening or narrowing disparities, providing a comprehensive equity diagnostic for academic outcomes and participation patterns.

C) Machine Learning Engine

6. Trend-Based Feature Construction and Model Training. For every SAT, ACT, and TSI metric, the system incorporates a campus's results from the previous one, two, and three years to capture historical performance patterns. Using these trend-informed features, three regression models were trained for each target variable: Linear Regression, Random Forest Regressor, and XGBoost Regressor (when available).

7. Next-Year Performance Forecasting. The trained models were used to generate predictions for the most recent year in the dataset. Forecasts were produced for SAT ERW, SAT Math, Total SAT, English ACT, Math ACT, and TSI readiness. All predicted values, alongside actual scores and model outputs, were stored across target-specific sheets for detailed campus-level comparison.

D) What-If Engine

8. Policy Simulation. The What-If Engine enables scenario testing by allowing users to modify key performance or participation variables (e.g., increasing SAT Math scores or raising participation rates). The system recomputes predictions, updated risk scores, and shifts in subgroup gaps, and also generates a narrative summary that interprets the projected impact.

E) Intervention Engine

9. Evidence-Based Recommendations. This module combines rule-based logic with an LLM-driven generation framework to produce targeted intervention suggestions. Recommendations are informed by risk indices, equity gaps, forecasted trends, and comparison benchmarks, and typically address areas such as math proficiency, college readiness, participation, and subgroup-specific supports.

4 Exploratory Data Analysis (EDA)

This section provides an overview of the combined SAT, ACT, and TSI dataset (2020–2023), with a focus on distributional patterns, yearly trends, subgroup differences, participation dynamics, and core correlations.

4.1 Core Distributions

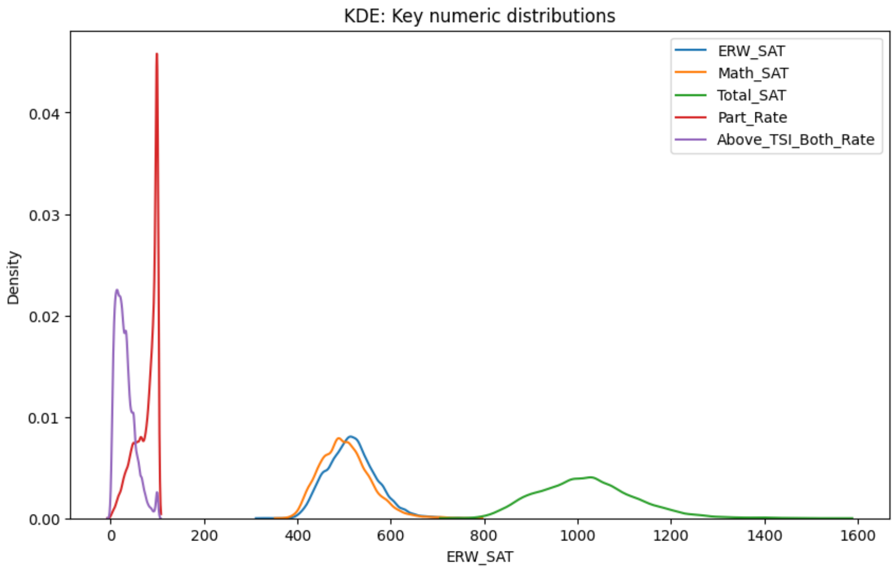

Figure 2: KDE distribution of core metrics (ERW, Math, Total SAT, Participation, TSI).

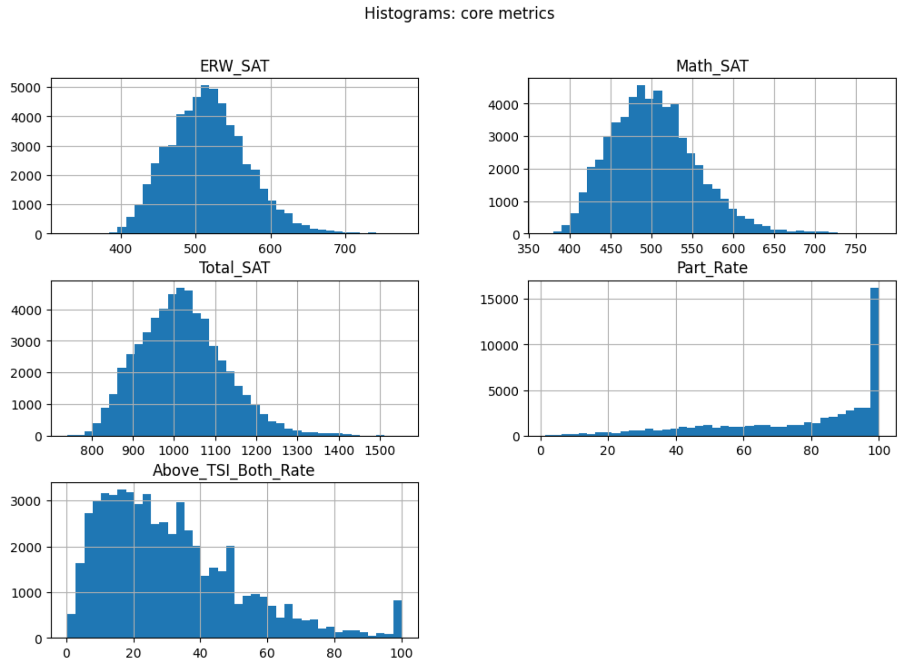

Figure 3: Histograms of core SAT and readiness metrics.

Observation: SAT Total scores show moderate right skew, while participation and TSI readiness have stronger left-skewed distributions, indicating concentrated low-readiness campuses.

4.2 Yearly Trends (2020–2023)

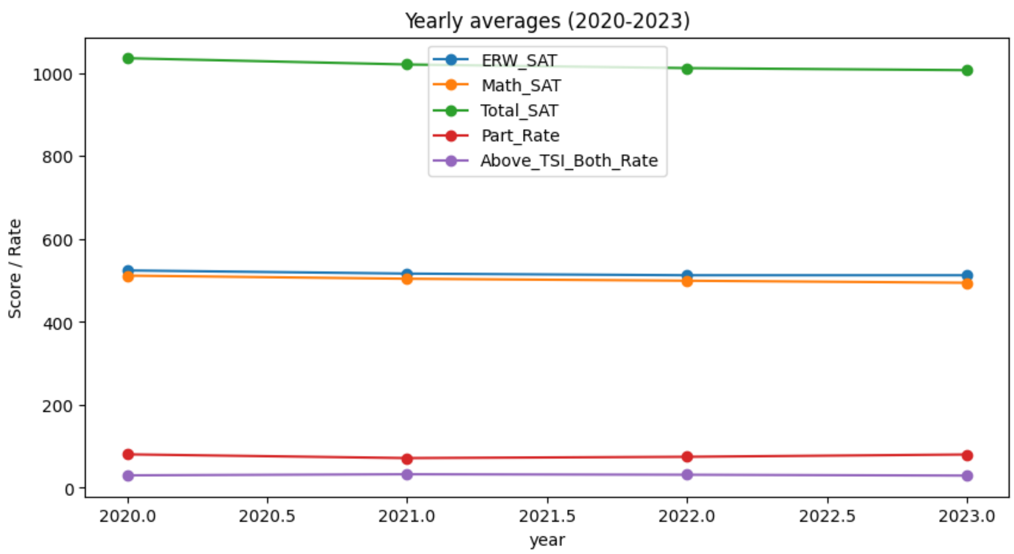

Figure 4: Year-wise mean ERW, Math, and Total SAT trends from 2020–2023.

Observation: Average SAT performance shows slight year-to-year fluctuation. Any dips or peaks may correspond to policy or participation shifts.

4.3 Group-wise Comparison

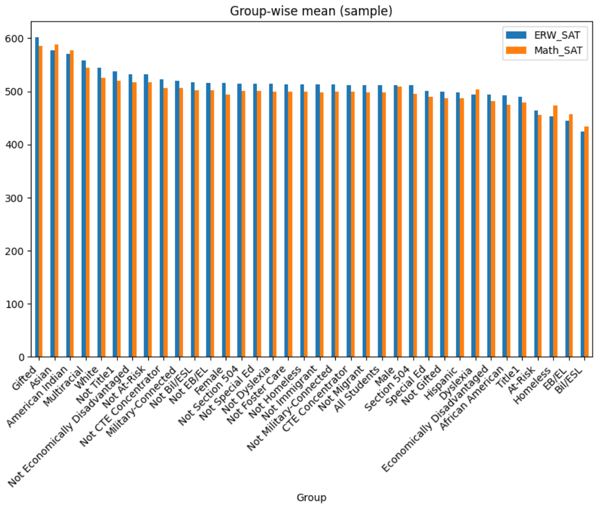

Figure 5: Group-wise mean scores for the latest available year.

Observation: Clear disparities exist among demographic groups, which are later quantified in the Equity Gap Analyzer.

4.4 Participation vs Performance

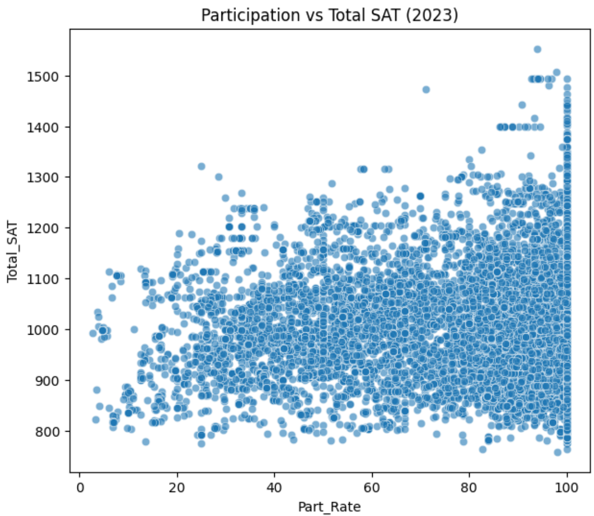

Figure 6: Scatterplot of participation rate vs Total SAT (2023).

Observation: A mild positive correlation is visible: campuses with higher participation often show stronger average SAT performance. This relationship informs our What-If Simulator.

4.5 Correlation Matrix

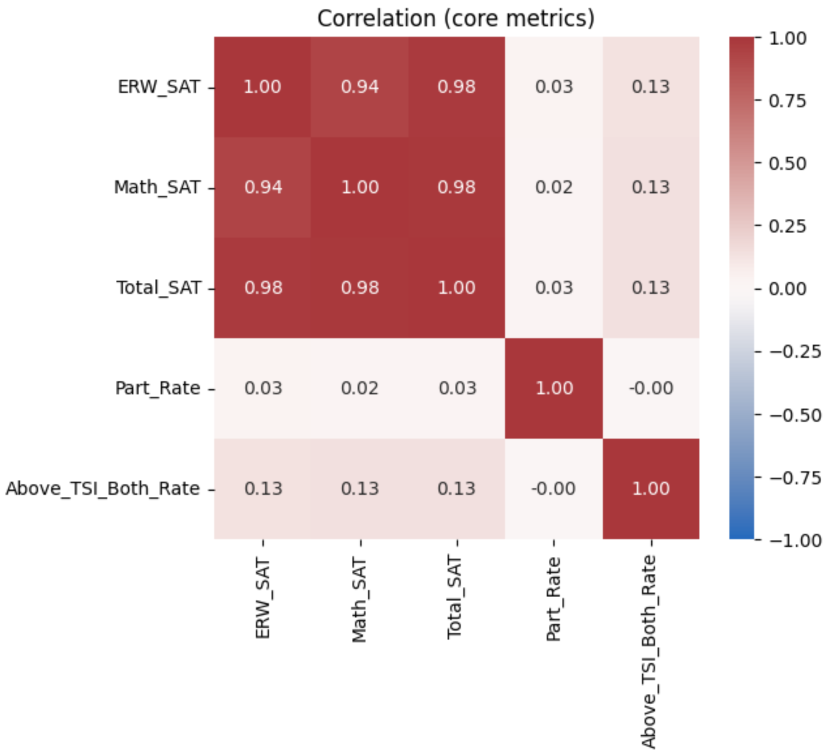

Figure 7: Correlation heatmap for ERW, Math, Total SAT, Participation, and TSI readiness.

Observation: ERW and Math correlate strongly with Total SAT, as expected. TSI readiness correlates moderately with core SAT components, while participation shows mixed correlations depending on campus context.

5 Summary of Outputs

Phase 1: Data Foundation Outputs

Phase 1 produced a cleaned, merged SAT–ACT–TSI dataset (2020–2023) and multiple diagnostic visualizations validating trends and group differences. The plots below were generated during Phase 1

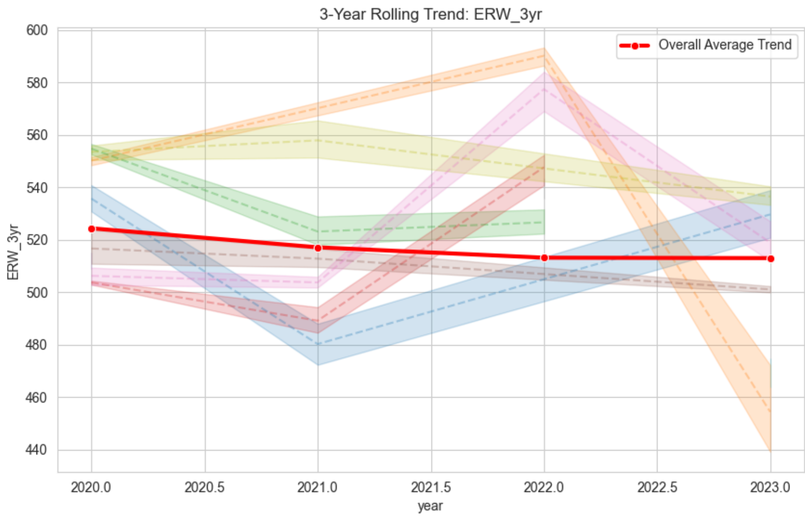

Figure 8: Three-year rolling trend for SAT ERW (Phase 1).

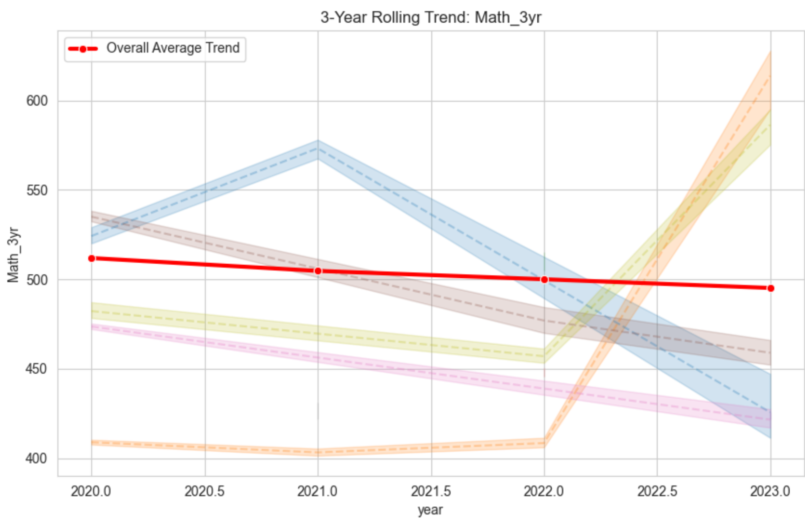

Figure 9: Three-year rolling trend for SAT Math (Phase 1).

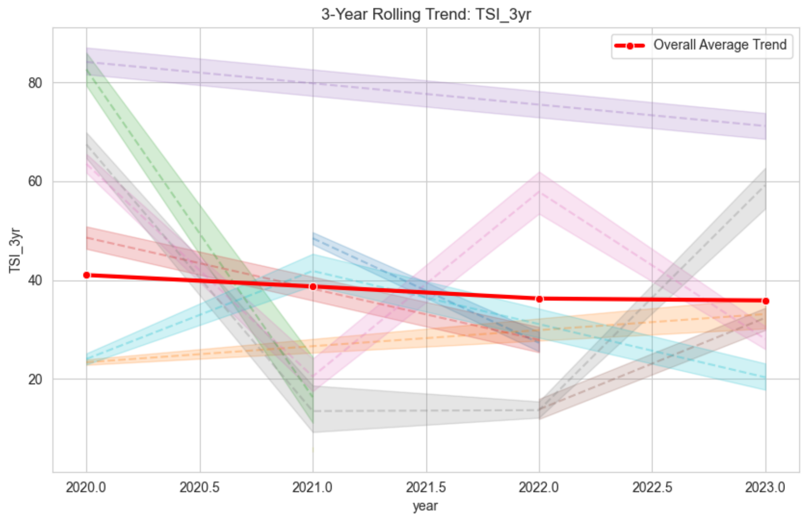

Figure 10: Three-year rolling trend for TSI readiness (Phase 1).

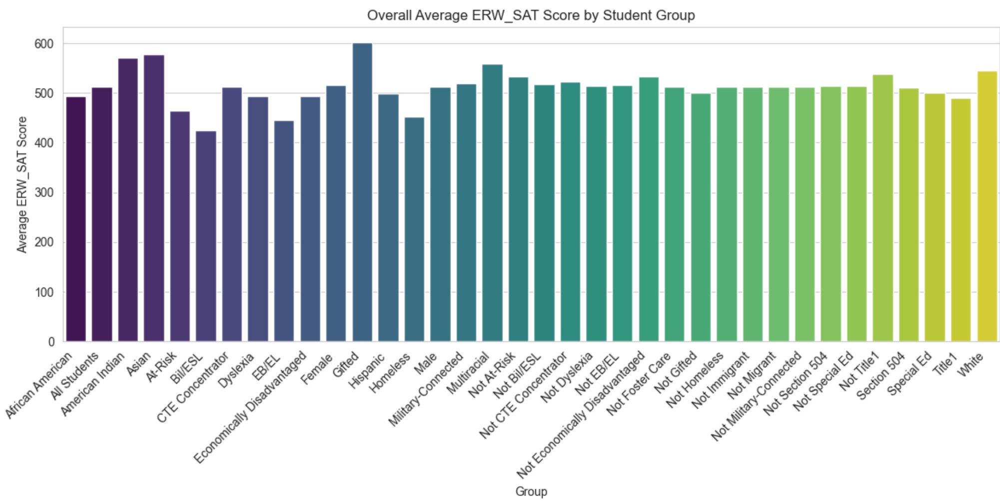

Figure 11: Group-wise comparison of ERW SAT averages (latest year snapshot).

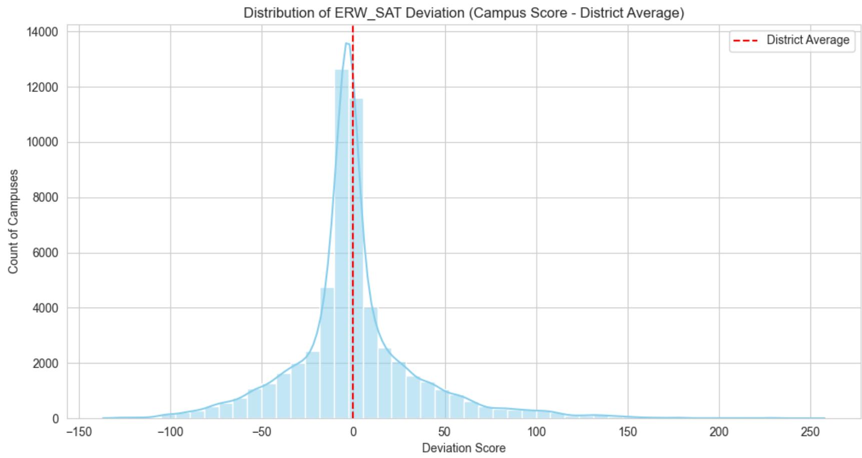

Figure 12: Campus vs. District deviation plot (Phase 1).

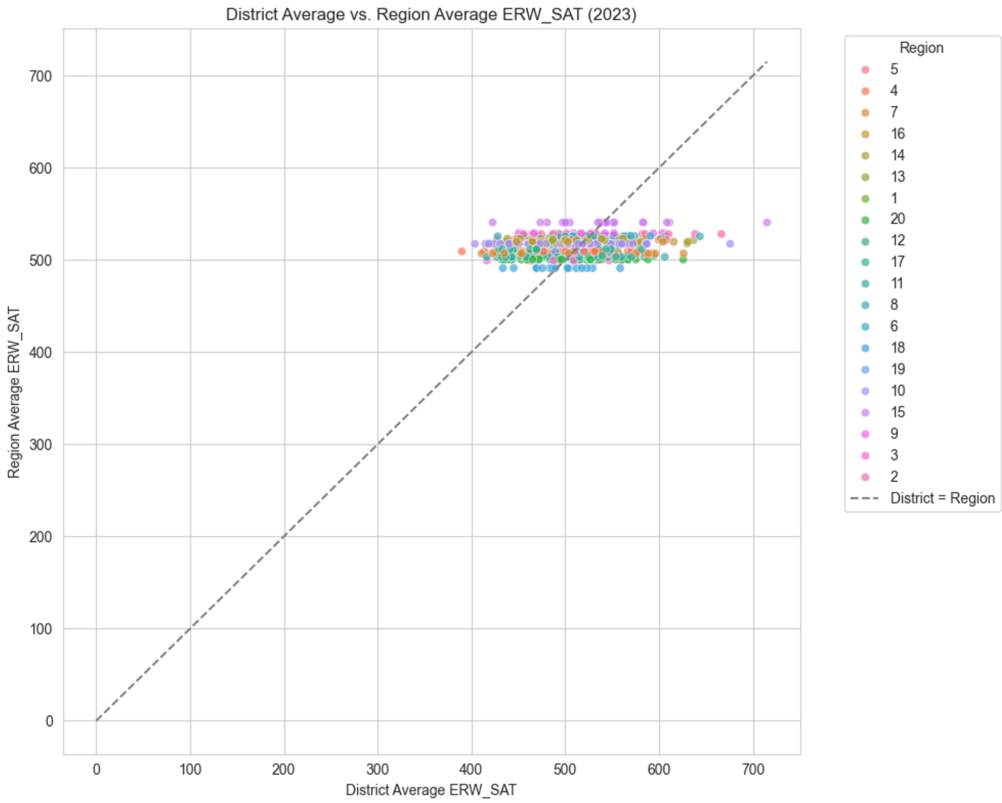

Figure 13: District vs. Region comparison scatter (Phase 1).

Summary: Phase 1 validated the merged dataset, produced multi-year trend signals for ERW, Math and TSI, and surfaced initial group- and geography-based deviations that feed directly into the Analytics Engine.

Phase 2: Analytics Engine Outputs

The Analytics Engine computed a composite Academic Risk Index (0–100) for each campus–year–group record. Components included SAT ERW and Math percentiles, TSI readiness, participation, multi-year decline, and group disparities. Phase 2 outputs include:

- `risk_score` column for all records,
- equity-gap tables comparing groups to campus and district benchmarks,

Phase 3: Machine Learning Outputs

Phase 3 developed a lag-based forecasting pipeline for all major SAT and ACT metrics (ERW_SAT, Math_SAT, English_ACT, Math_ACT, Total_SAT). Using 1-, 2-, and 3-year campus-level lag features, three models were trained for each target: **Linear Regression**, **Random Forest**, and **XGBoost**. Model performance was evaluated using MAE, RMSE, R^2, adjusted R^2, and prediction accuracy (percentage within ± 20 points of actual values). All model artifacts were saved, and predictions and metrics were exported

Table 1: Phase 3 Model Performance Summary

Target	**Model**	**MAE**	**RMSE**	R^2	**Adj.** R^2	**Accuracy** ($\pm$20)
ERW_SAT	Linear Regression	21.005	33.021	0.644	0.643	68.05%
	Random Forest	19.149	30.507	0.696	0.696	70.72%
	XGBoost	19.463	31.606	0.674	0.673	70.07%
Math_SAT	Linear Regression	20.437	32.527	0.668	0.668	69.16%
	Random Forest	18.622	30.115	0.715	0.715	71.77%
	XGBoost	19.167	31.831	0.682	0.682	70.64%
English_ACT	Linear Regression	1.279	2.087	0.694	0.694	100.0%
	Random Forest	1.284	2.038	0.708	0.708	100.0%
	XGBoost	1.284	2.110	0.688	0.687	100.0%
Math_ACT	Linear Regression	1.053	1.713	0.681	0.681	100.0%
	Random Forest	1.007	1.634	0.710	0.709	100.0%
	XGBoost	1.008	1.686	0.691	0.690	100.0%
Total_SAT	Linear Regression	40.916	64.668	0.656	0.656	42.96%
	Random Forest	37.362	59.794	0.706	0.705	47.43%
	XGBoost	37.776	61.484	0.689	0.689	48.90%

Phase 4: What-If Engine (LLM-Enabled Simulation)

The What-If Engine is an LLM-enabled agent that allows a user to **select any campus, any group, and any year**, change inputs (for example: participation, SAT Math or

ERW, TSI rates), and instantly obtain updated predicted outcomes and an LLM-crafted narrative explaining the projected impact.

Illustrative scenario (Campus 441165, Year 2022):

- SAT Participation: **63.0%** → **90.0%**
- SAT Math: **496** → **516** (+20 points)

Table 2: What-If results (Campus 441165, 2022)

Metric	Baseline	Scenario
SAT Participation Rate (%)	63.0	90.0
SAT Math	496	516
Predicted Total SAT	1009	1029
TSI Readiness (%)	27.6	27.6
Academic Risk Score	40.72	40.72

Interpretation: The scenario yields a measurable increase in predicted Total SAT but no change in the aggregated Risk Score, indicating persistent subgroup-level risk drivers that broad participation or average score increases do not resolve.

Phase 5: Intervention Engine (LLM-Based Recommendations)

The Intervention Engine leverages the Analytics Engine outputs, ML forecasts, and What-If simulations to generate prioritized, evidence-based interventions via an LLM agent. Recommendations are contextual and change depending on the selected campus/year/scenario.

Example recommendations for Campus 441165 (2022):

- **Low TSI readiness (27.6%)**: Implement mandatory credit-bearing TSI preparation courses for juniors and seniors below readiness benchmarks. Expected effect: ∼+15 percentage points in one year.
- **Below-average SAT Math**: Launch targeted after-school SAT Math preparation focused on algebra and geometry fundamentals. Expected effect: ∼+20 SAT points for participating students.

Note: Because both What-If and Intervention agents are LLM-driven, users can request natural-language explanations, multi-step intervention plans, and monitoring metrics tailored to any campus or subgroup.

Dashboard

A lightweight dashboard (Streamlit prototype) was built to demonstrate interactive functionality. Dashboard features include:

- Campus overview and subgroup breakdowns,
- Academic Risk Index explorer,
- Benchmarks (Campus vs District vs Region),
- Interactive What-If sliders (to modify participation and score inputs),
- Real-time LLM-generated intervention recommendations.

Concluding remark: These outputs collectively form a practical educational intelligence platform that supports data-driven identification of at-risk cohorts, simulation of policy levers, and generation of targeted interventions suitable for operational use by district leaders.

6 Discussion

The multi-agent educational intelligence framework developed in this project demonstrates how traditional analytics, machine learning, and LLM-based reasoning can be integrated into a unified decision-support system for academic performance monitoring.

Data Quality and Multi-Year Trends

The merged SAT–ACT–TSI dataset (2020–2023) revealed several structural patterns. Multi-year trends showed moderate fluctuations in SAT ERW and Math performance, with noticeable variations across campuses and student groups. Participation patterns varied widely, creating differences in observed averages and making lag-based modeling essential. These Phase 1 and Phase 2 findings underscored the need for a risk-sensitive framework that contextualizes performance across time, geography, and subgroup composition.

Interpretation of the Academic Risk Index

The Academic Risk Index generated in Phase 2 proved effective in summarizing multi-dimensional academic signals into an interpretable 0–100 scale. The index consistently highlighted campuses with low SAT/ACT proficiency, declining multi-year performance, and substantial subgroup disparities. Importantly, the index did not always move in parallel with score improvements as observed in the What-If scenario indicating that structural equity gaps remain even when averages rise. This reinforces the value of a composite, trend-aware metric rather than relying solely on annual performance.

Model Performance and Forecast Reliability

Phase 3 regression models (Linear Regression, Random Forest, XGBoost) produced stable forecasts using campus-level lag features. The inclusion of 1-, 2-, and 3-year lag variables substantially improved predictive stability. Evaluation metrics (MAE, RMSE, R^2, adjusted R^2, and accuracy within ± 20 points) indicate that the models are suitable for high-level policy forecasting, though not intended as student-level predictors. Forecast reliability varied by target variable, with Total SAT and component SAT metrics generally more predictable than TSI readiness, which is inherently more volatile and sensitive to policy shifts.

Insights from the What-If Engine

Phase 4 demonstrated the practical value of scenario simulation. The example scenario for Campus 441165 illustrated that substantial increases in participation and SAT Math can improve predicted outcomes without necessarily reducing overall risk. This suggests that broad interventions (e.g., improving instruction or boosting participation) may not fully address subgroup-based disparities. The ability for users to select any campus, year, or group and generate customized scenarios makes the What-If Engine a highly flexible tool for leadership planning.

Effectiveness of LLM-Driven Interventions

The Intervention Engine (Phase 5) effectively transformed numerical insights into actionable strategies tailored to each campus. The LLM-generated recommendations were aligned with campus-specific weaknesses such as low TSI readiness or weak Math performance, and they offered feasible interventions (TSI bootcamps, SAT Math tutoring, targeted academic supports). Because these recommendations adapt to user-selected scenarios, the system supports iterative decision-making and helps leaders understand both immediate and structural needs.

Usability and Practical Deployment

The lightweight dashboard integrates all four agents Analytics, Prediction, What-If, and Intervention into an accessible interface. By enabling real-time scenario adjustments, quick visualizations, and natural-language explanations, the dashboard lowers technical barriers for district administrators. While the current prototype focuses on SAT/ACT/TSI datasets, the architecture is extensible to additional academic and behavioral indicators.

Limitations and Future Opportunities

Several limitations deserve recognition. First, lag-based forecasting assumes stability in assessment practices, which may not hold during policy changes. Second, masked data (<25) and heterogeneous reporting standards introduce noise into subgroup analyses. Third, the Risk Index, though comprehensive, is sensitive to the chosen weights. Future work may incorporate Bayesian updating, causal modeling, or reinforcement learning to simulate longer-term impacts of interventions. Expanding the dashboard to include student-level drilldowns, attendance, and course completion metrics would further enhance its operational usefulness.

Overall Contribution

Overall, the system provides a novel, integrated approach to academic intelligence by combining data engineering, analytics, forecasting, simulation, and LLM reasoning. It supports proactive planning, equitable resource allocation, and evidence-based leadership decisions, marking a significant advancement over traditional static reporting systems.

7 Future Work

Looking ahead, the system can be expanded into a scalable academic intelligence framework capable of supporting multiple data domains, larger districts, and more complex decision workflows. The key directions for future development are outlined below:

1. **Generalization to New Assessments and States.** The architecture can be extended beyond SAT–ACT–TSI to incorporate STAAR, AP, dual-credit, attendance, or graduation data. Because the agents operate on standardized feature schemas, adding new domains requires minimal modification, enabling deployment across different states or districts.

2. **Model Scalability and Distributed Training.** As districts scale to hundreds of campuses and multi-year histories, training can be shifted to distributed environments (e.g., Spark, Ray, or cloud ML pipelines). This would allow parallel model training for each subject, campus cluster, or demographic segment while maintaining low inference latency for the What-If Engine.

3. **Unified Feature Store for Reusability.** A centralized feature store (e.g., Feast) could standardize lag features, trend calculations, and risk indicators across agents. This ensures reproducibility, version control, and seamless re-use of engineered features across analytics, forecasting, simulation, and intervention modules.

4. **Dynamic Scenario Optimization.** The current What-If Engine simulates user-defined scenarios. A scalable extension would automatically search the intervention space (participation boosts, remediation strategies, subgroup targeting) to identify *optimal* policy levers under constraints such as cost, staffing, or time.

5. **Cross-Domain Transfer and Adaptive Models.** By incorporating domain adaptation or meta-learning, the forecasting agents could learn shared patterns across campuses with similar demographic or performance profiles. This improves accuracy for small or data-sparse schools and makes the system more resilient across districts.

6. **Modular Multi-Agent Expansion.** The existing four-agent structure can evolve into a modular suite (EquityAgent, ResourceAgent, AttendanceAgent, EarlyWarningAgent). Each agent would subscribe to the shared feature store, allowing the platform to grow horizontally without rewriting core logic.

These improvements emphasize scalability, cross-domain applicability, and architectural extensibility, positioning the platform as a long-term, district-wide decision-support system rather than a single-use analytical tool.

8 Conclusion

This work presents a scalable, data-driven academic intelligence system that integrates longitudinal analytics, machine learning forecasting, scenario modeling, and LLM-guided intervention design. By unifying SAT, ACT, and TSI readiness data from 2020–2023 into a single standardized panel, the system enables consistent comparisons across campuses, groups, and years. The Analytics Engine surfaces structural patterns such as declining multi-year trends and persistent subgroup disparities while the Academic Risk Index provides a stable, interpretable indicator of overall readiness.

The forecasting results demonstrate that lag-based models are effective for campus-level prediction. Random Forest and XGBoost models achieved competitive performance across key metrics, with R^2 values between 0.68 and 0.72 for SAT components and prediction accuracies exceeding 70% within ±20 points. ACT English and Math forecasts were particularly stable, with all models achieving 100% accuracy within the same tolerance band. These results confirm that historical campus-level performance contains sufficient signal for reliable trend estimation and medium-term academic planning.

The What-If Engine and Intervention Agent translate these analytic outputs into practical decision-support tools. Scenario simulations reveal how changes in participation, instruction, or subgroup performance influence overall readiness, while LLM-generated recommendations provide administrators with targeted, context-sensitive strategies. Together, these components move beyond descriptive reporting and enable forward-looking, interactive planning aligned with district priorities.

Overall, the system demonstrates that combining structured analytics with modern machine learning and generative models can produce a robust, extensible platform for academic decision-making. It not only identifies existing performance gaps but also forecasts future risks, tests policy changes, and recommends actionable interventions. With further scaling and integration, the platform has the potential to function as a district-wide intelligence layer supporting continuous improvement, equitable resource allocation, and data-driven leadership at scale.

References

[1] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," *IEEE Transactions on Systems, Man, and Cybernetics*, 2010.

[2] M. Yağcı *et al.*, "Predicting student performance using machine learning," *Education Sciences*, 2022.

[3] "Performance prediction using educational data mining," *Scientific African*, 2025.

[4] "Explainable student performance prediction models," *Journal of Information Systems Education*, 2025.

[5] N. S. Hassan *et al.*, "Predicting students' final gpa using decision trees: A case study," *International Journal of Information and Education Technology*, 2016.

[6] "Machine learning in educational science: Opportunities and challenges," *Review of Education*, 2023.

[7] "Llm agents in education: A systematic review," *Computers and Education: Artificial Intelligence*, 2025.

[8] "Llm-based multi-agent systems in education," in *Advances in AI for Education*, 2024.