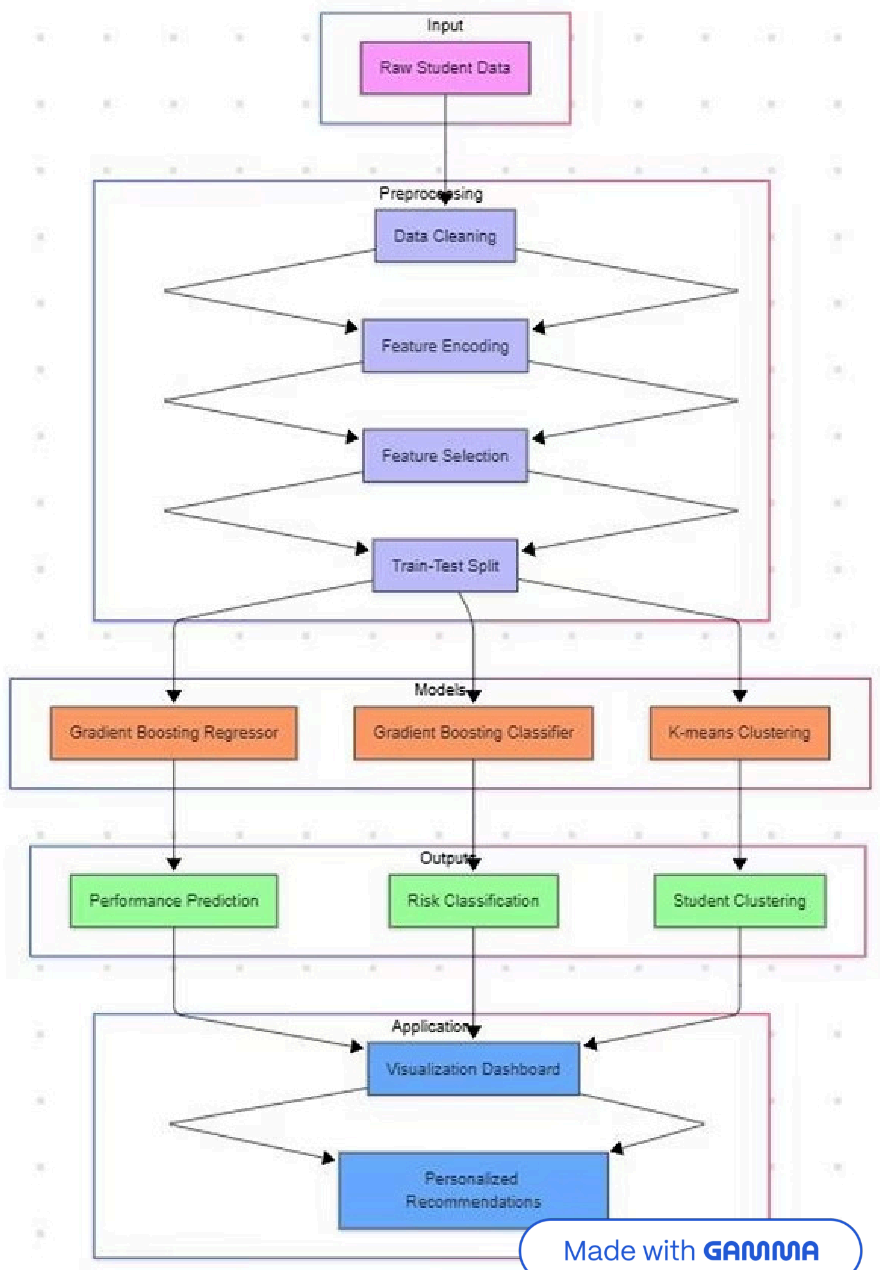
ACHIEVE - Academic Child Performance Evaluation & Recommendation System

ABSTRACT :

Machine learning predicts student success

Gradient Boost & K-means: $R^2 = 0.84$, Accuracy = 86%

Key factors: parental education, test prep, SES



Why ACHIEVE?

Traditional Assessments Fail

Miss early risk indicators

Data Science Bridge

Connects gaps in educational evaluation

Smart Goals

Identify predictors, build models, classify needs

Literature Review

Model Superiority

Gradient Boosting outperforms traditional regression

Optimal Clustering

3-5 student clusters work best (K-means)

Key Features

Parental education, SES, test prep most important

Ethical Framework

DELICATE framework guides responsible use

Comparative Study

Study	Method	Size	Accuracy	Limitation
Kotsiantis	Decision Tree	300	78%	Small sample
Chen	XGBoost	15,000	R^2=0.89	Limited demographics
Hämäläinen	K-means	2,400	4 clusters	No longitudinal data
Ours	Ensemble	1,000	R^2=0.84, 86%	One institution

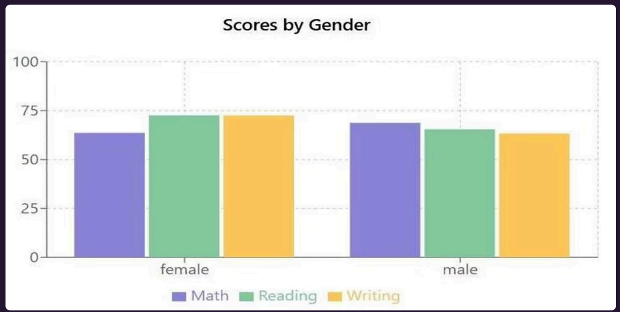
Dataset & Preprocessing

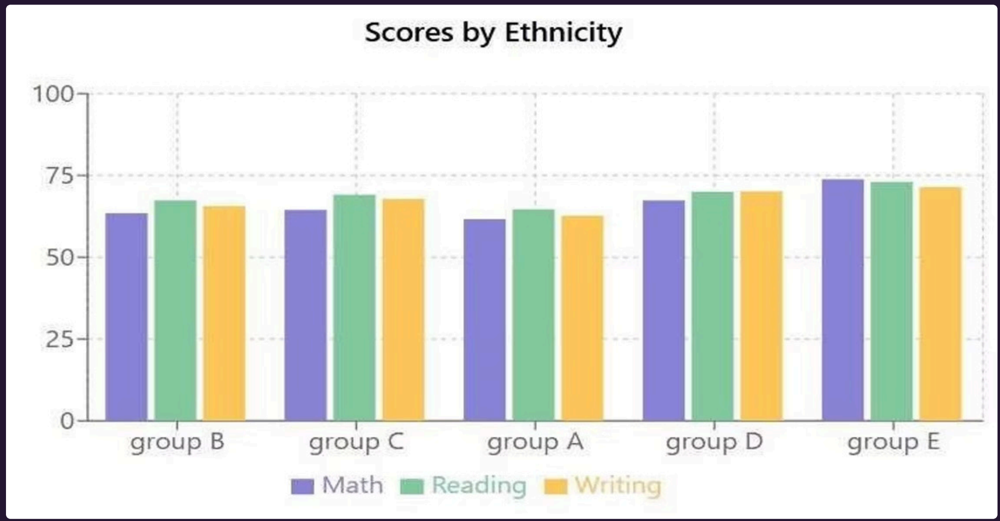
Data Details

- 1,000 students
- Demographics: gender, ethnicity
- Background: parental education, test prep
- Scores: math, reading, writing

Preprocessing Steps

- Cleaning, encoding, scaling
- Feature engineering
- PCA → 5 features retained





```
Accuracy: 0.88
Precision: 0.71
Recall: 0.19
```

Modeling Techniques

0.71

Linear

R^2 score

0.84

Random Forest

R^2 score

0.86

Gradient Boosting

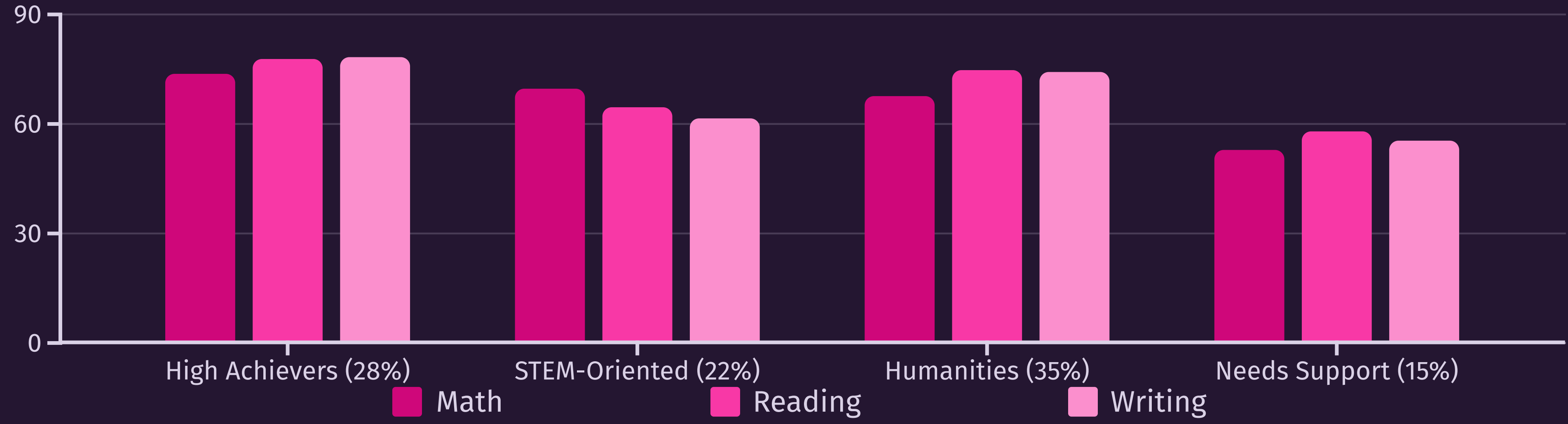
Best accuracy

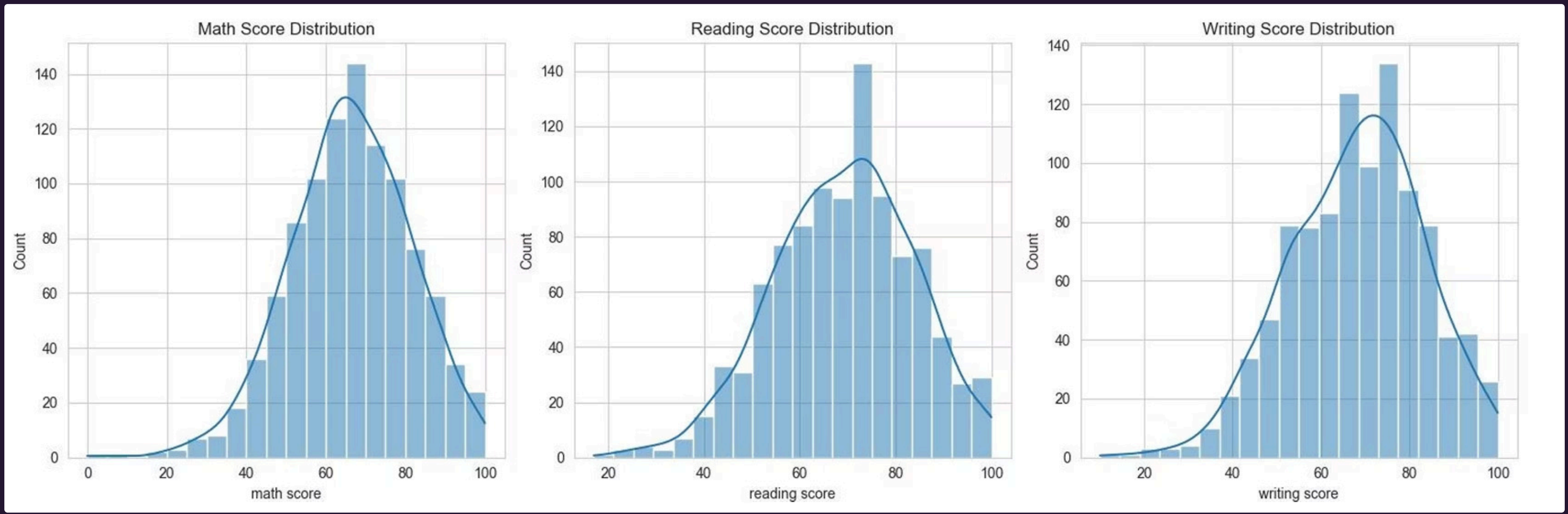
0.85

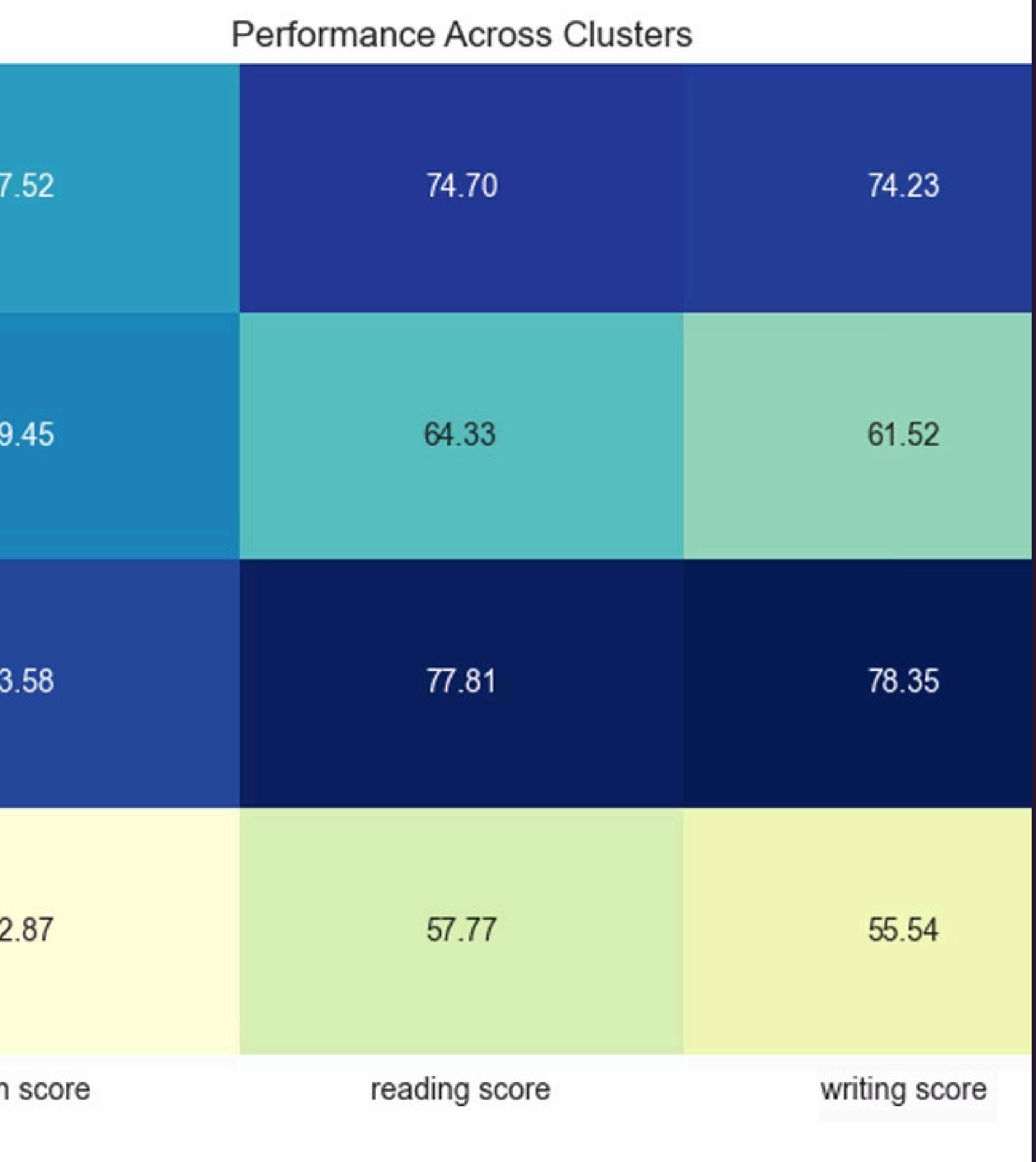
F1 Score

Classification performance

Clustering Results







Results & Insights

Gender Gap

Girls scored higher in reading/writing

Ethnicity Impact

Group E > Group A by 9.76 points

Lunch Status

Standard lunch students scored higher

Test Prep Boost

9.92 point increase in writing scores

Dashboard + Ethics

Visualize Trends

Track performance patterns over time

Detect At-Risk

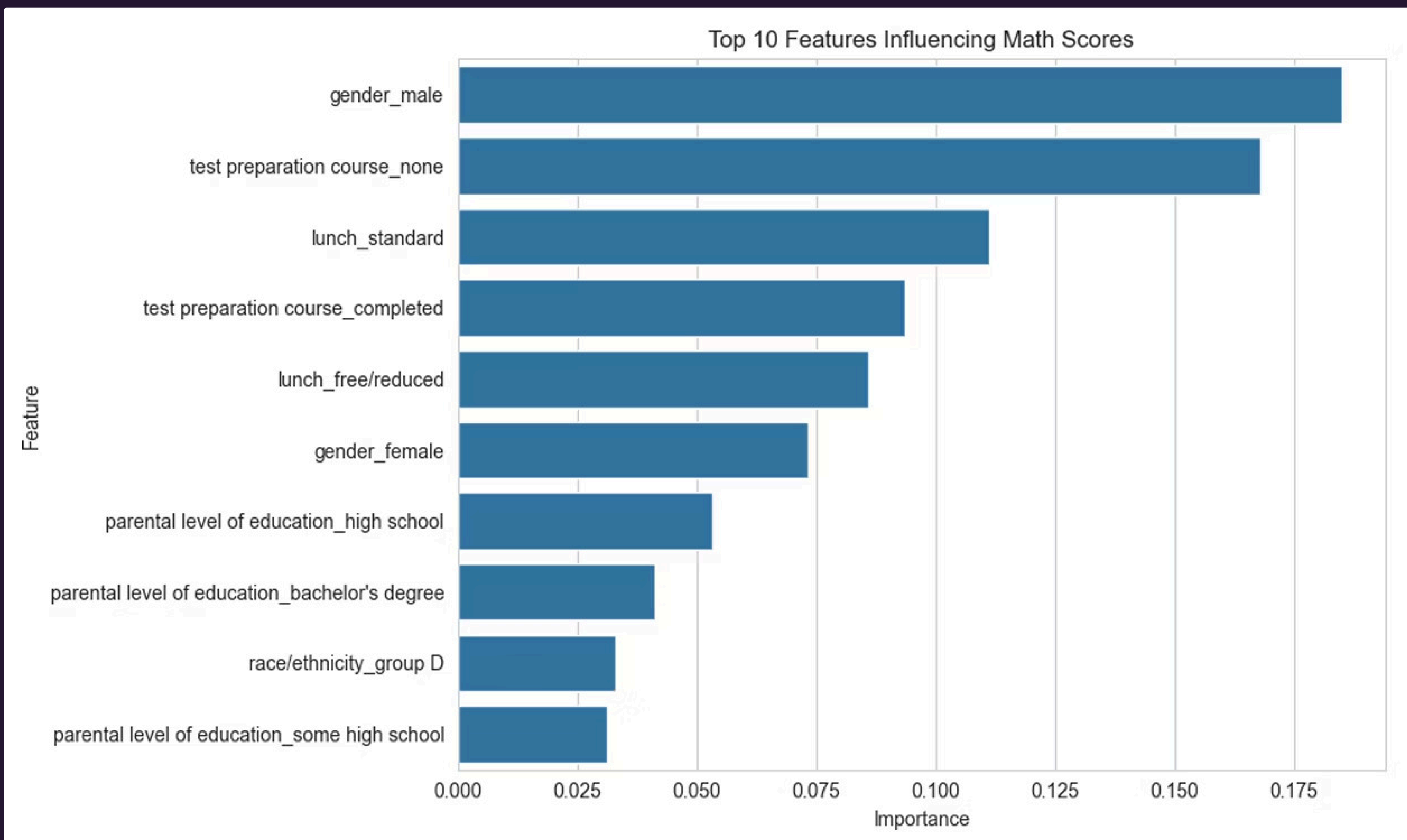
Early identification of struggling students

Ethical Design

Gender bias controlled, differential privacy

Evaluate Interventions

Measure effectiveness of support programs



Conclusion & Future Work

Achievements

Accurate predictions, effective clustering

Track Growth

Monitor student progress over time

Expand Data

Include attendance, co-curriculars

Mobile App

Launch parent interface

Mining Patterns and Clusters from Student Scores

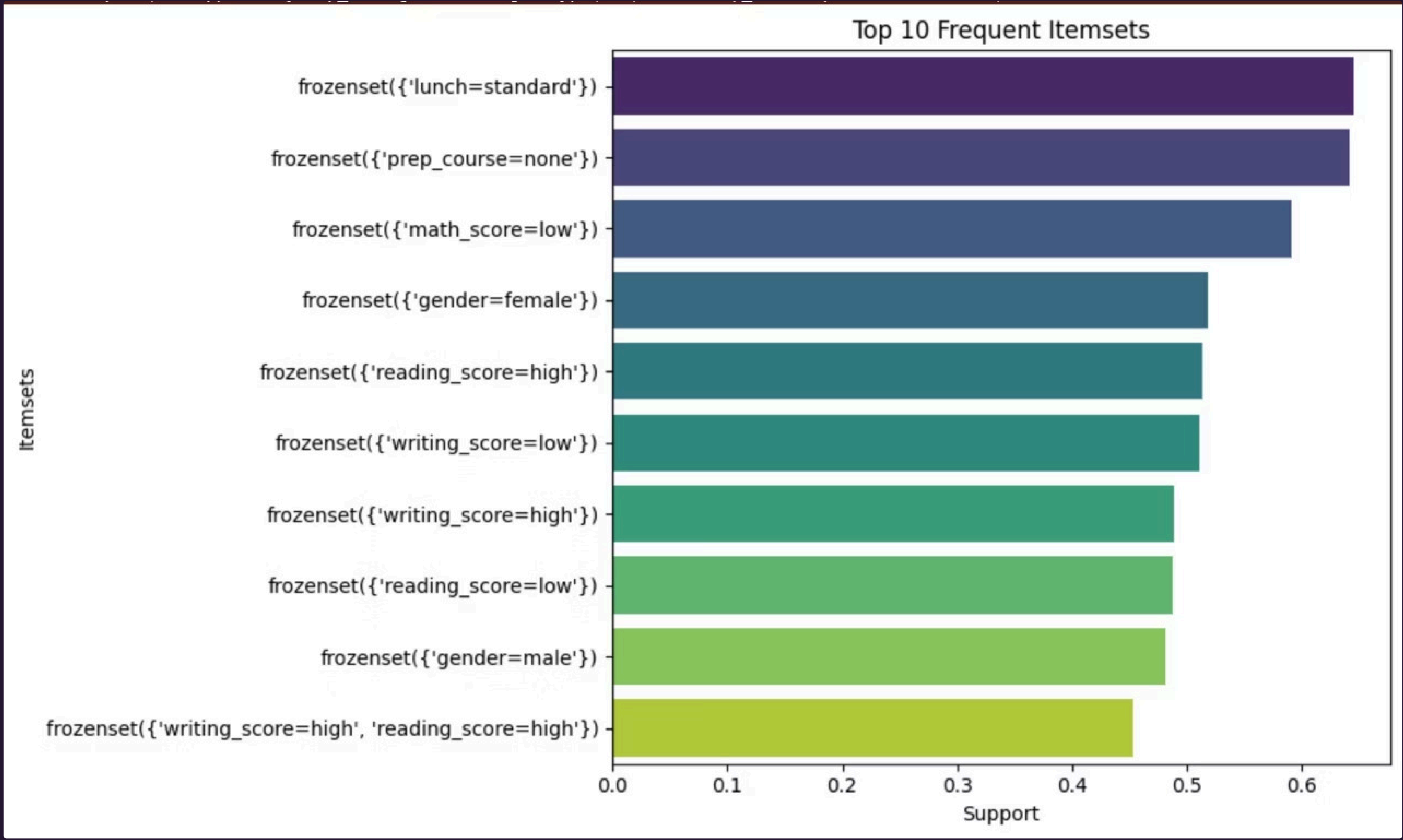
Association Rule Mining

- **Apriori:** Finds frequent itemsets with strong rules
- **FP-Growth:** Faster pattern mining alternative
- Results show links: test prep → high reading/writing scores
- Parental education strongly correlates with performance

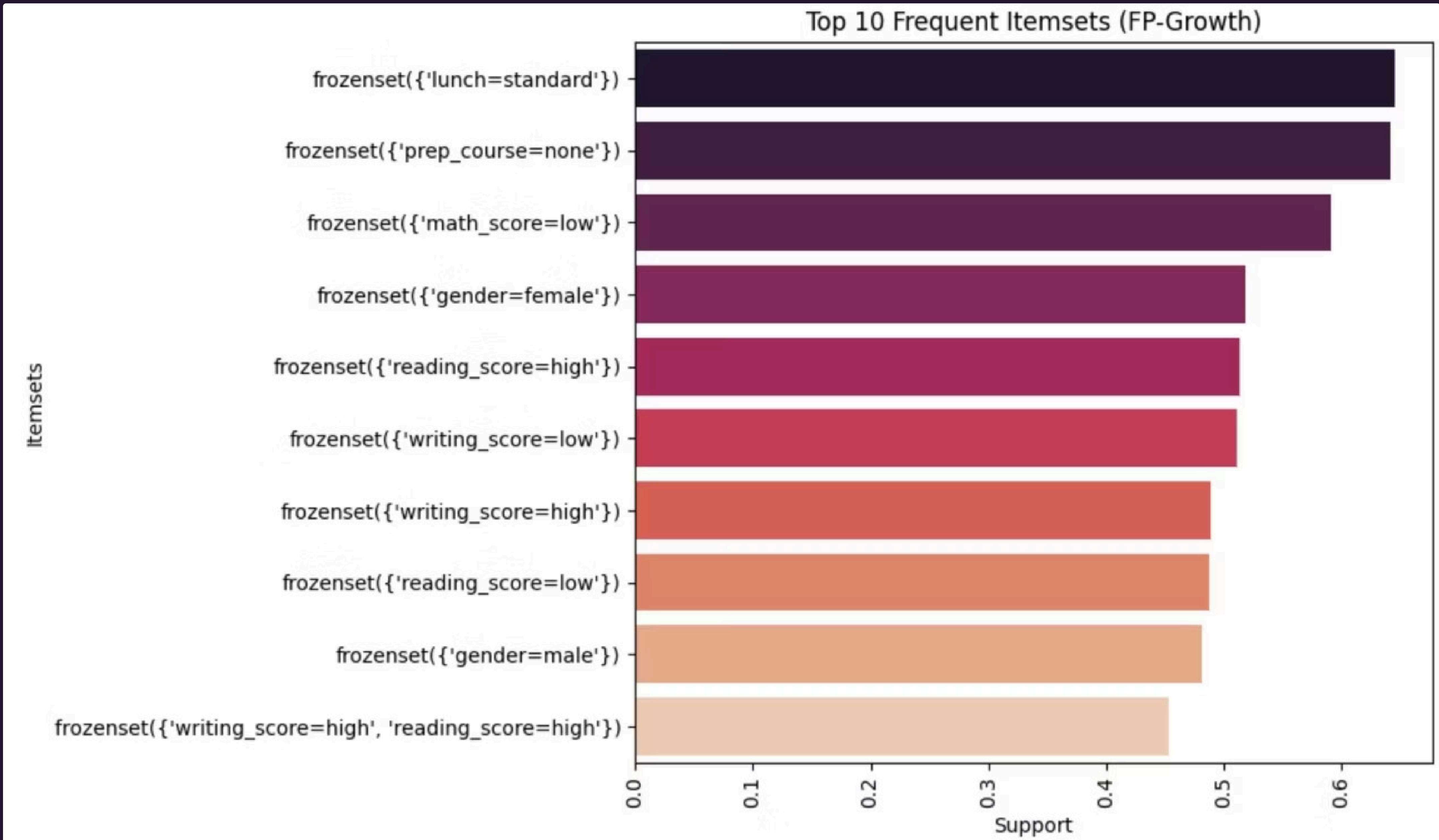
DBSCAN Clustering

- Clusters students by math, reading, writing scores
- No need for predefined labels
- Visualized with 3D PCA scatter and 2D score plots
- Reveals natural performance groups for targeted support

Output of Apriori Algorithm:



Output of FP Growth Algorithm:



The Results:

```
Top 5 Association Rules:
                                      antecedents  \
697          (writing_score=high, gender=male)
698                          (math_score=high)
413  (writing_score=high, reading_score=high)
416           (math_score=high, gender=female)
467                          (math_score=high)

                                     consequents  support  confidence      lift
697                          (math_score=high)    0.154    0.939024  2.295903
698          (writing_score=high, gender=male)    0.154    0.376528  2.295903
413           (math_score=high, gender=female)    0.176    0.388521  2.182702
416  (writing_score=high, reading_score=high)    0.176    0.988764  2.182702
467          (reading_score=high, gender=male)    0.173    0.422983  2.169143
```

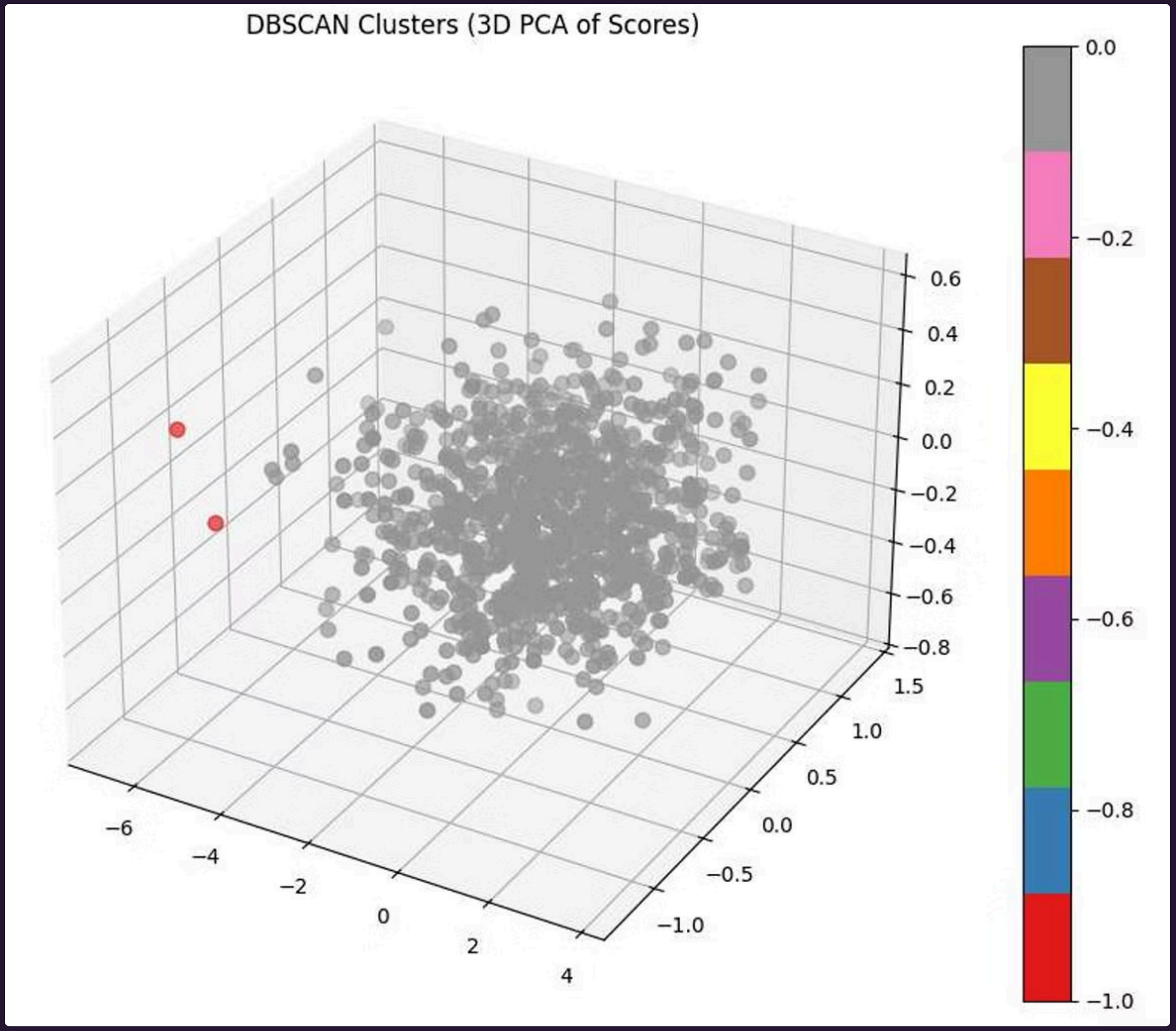
Apriori Algorithm

Top 5 FP-Growth Association Rules:

	antecedents	consequents	support	confidence	lift
684	(math_score=high)	(writing_score=high, gender=male)	0.154	0.376528	2.295903
683	(writing_score=high, gender=male)	(math_score=high)	0.154	0.939024	2.295903
417	(writing_score=high, reading_score=high)	(math_score=high, gender=female)	0.176	0.388521	2.182702
420	(math_score=high, gender=female)	(writing_score=high, reading_score=high)	0.176	0.988764	2.182702
466	(reading_score=high, gender=male)	(math_score=high)	0.173	0.887179	2.169143

FP Growth Algorithm

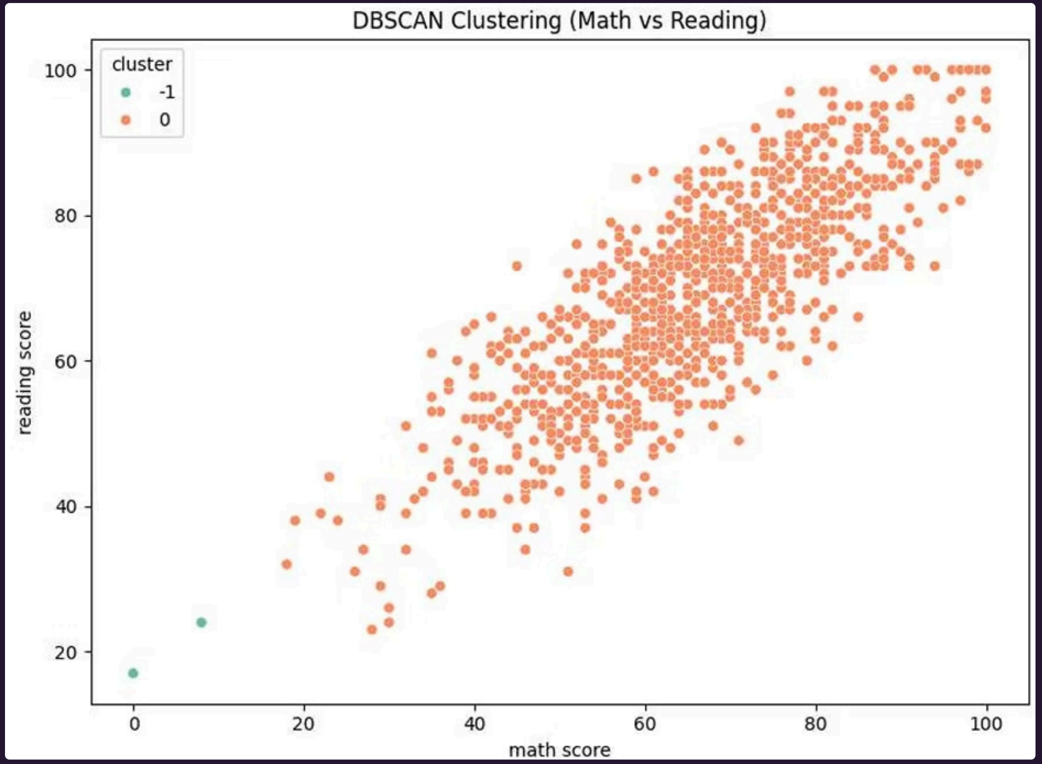
DB-SCAN Algorithm:

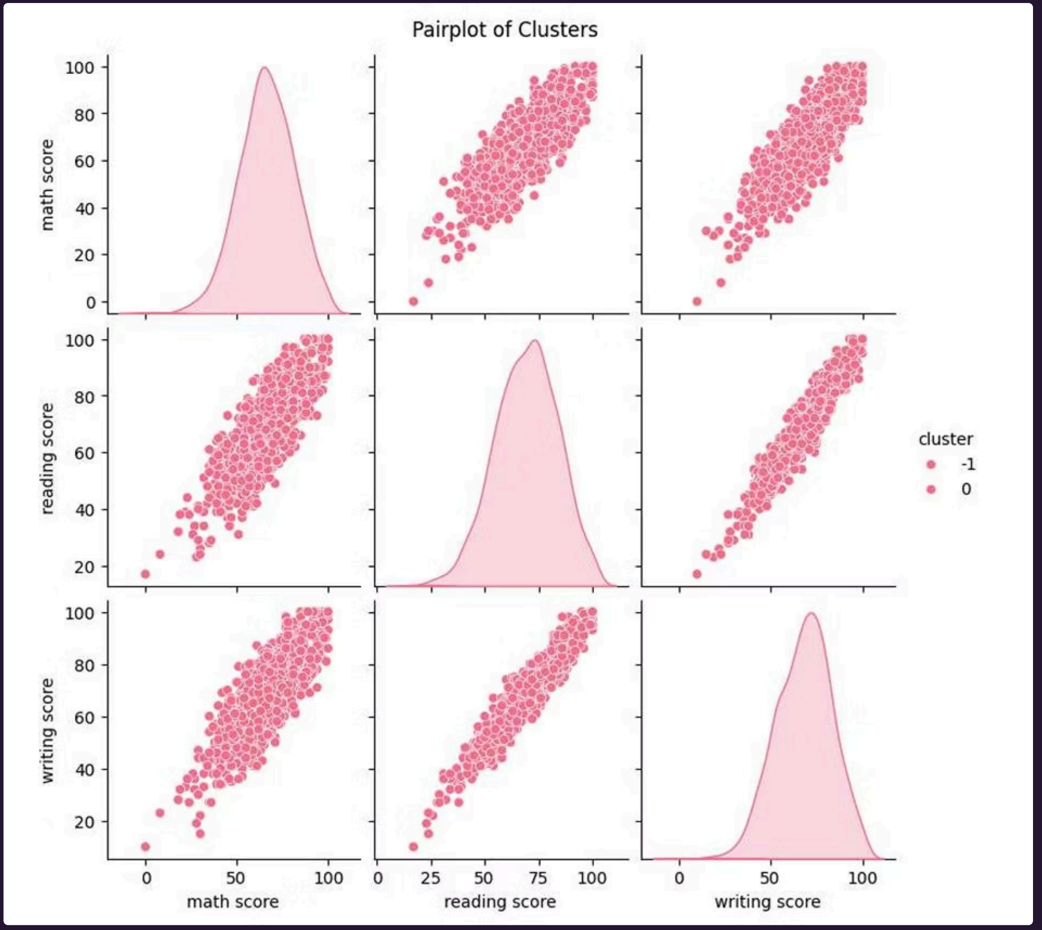


The 3D PCA Scatter Plot visualizes the clusters formed by the DBSCAN algorithm based on students' **math**, **reading**, and **writing scores**.

- **PCA (Principal Component Analysis)** reduces the 3D score data to a new 3D space that captures the most significant variance in performance.
- Each point represents a student, and the **color indicates the cluster** assigned by DBSCAN.
- DBSCAN groups students with similar academic performance into distinct clusters while **automatically identifying noise/outliers**, which appear as separate or unclustered points.
- The clear separation between clusters in this plot reveals how performance patterns naturally group — without needing to predefine the number of clusters like in K-Means.

This plot helps us visually interpret how students with similar strengths or weaknesses in scores tend to cluster together, and highlights any students who deviate significantly from the norm.





Made with GAMMA

Top 10 Features by SHAP Value

Parental Education (Master's)

Highest influence (0.32 SHAP value)

Test Preparation Completion

Significant impact (0.29 SHAP value)

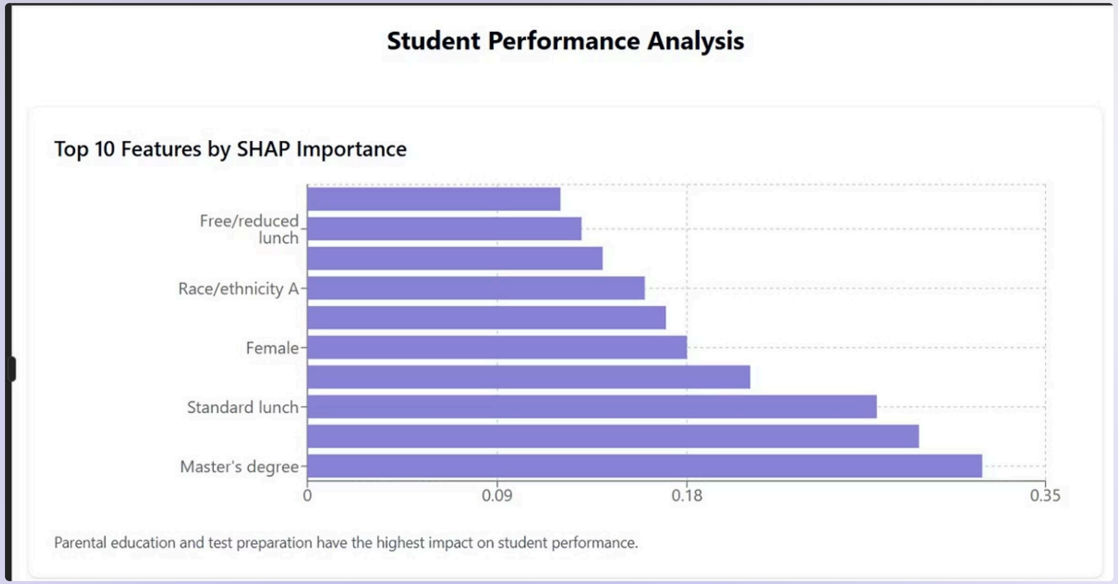
Standard Lunch Type

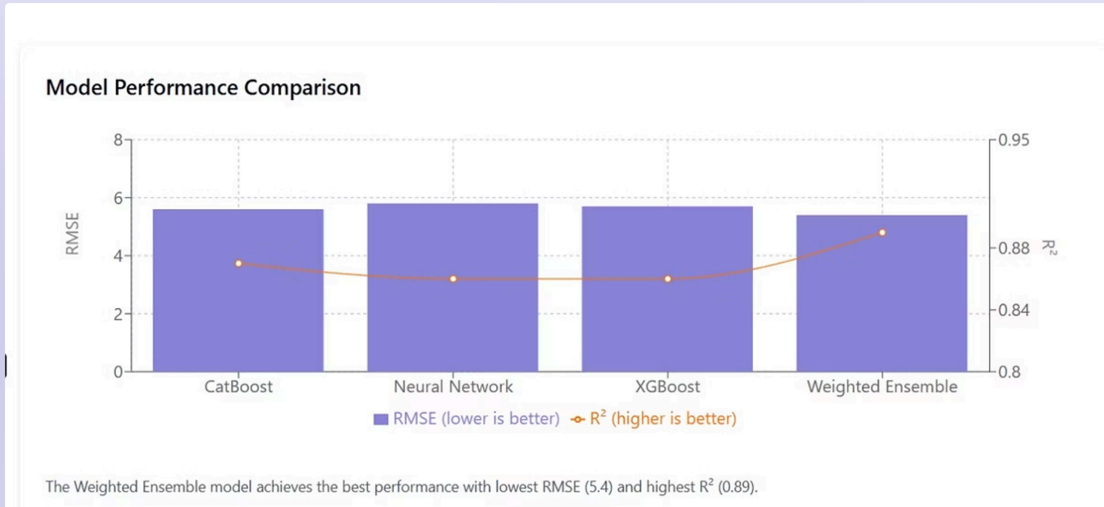
Strong socioeconomic factor (0.27 SHAP value)

Race/Ethnicity & Gender

Measurable roles in performance

Home education and resources are key drivers of success.





Predictive Model Accuracy Comparison

Weighted Ensemble Model

Best performance: RMSE 5.4, R^2 0.89

Individual Models

CatBoost, Neural Network, XGBoost strong but lower accuracy

Ensemble Approach

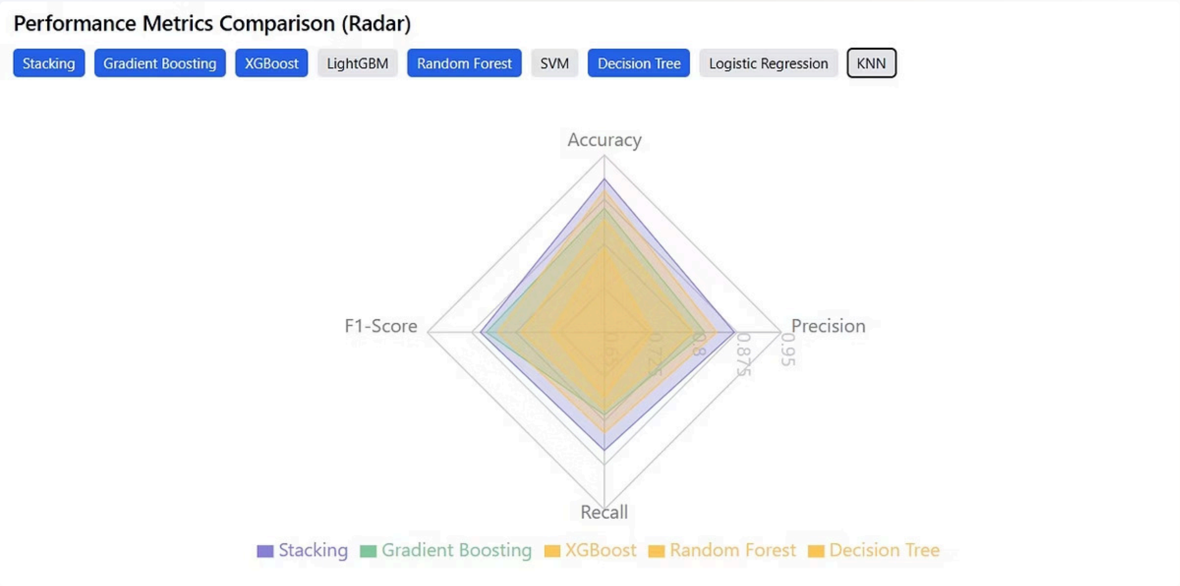
Combines strengths, improves prediction accuracy

Ensembles yield more reliable student performance forecasts.

Conclusions and Recommendations

1. Home education strongest success predictor
2. Expand test preparation programs
3. Address socioeconomic disparities
4. Develop gender-specific learning approaches
5. Use combined models for prediction

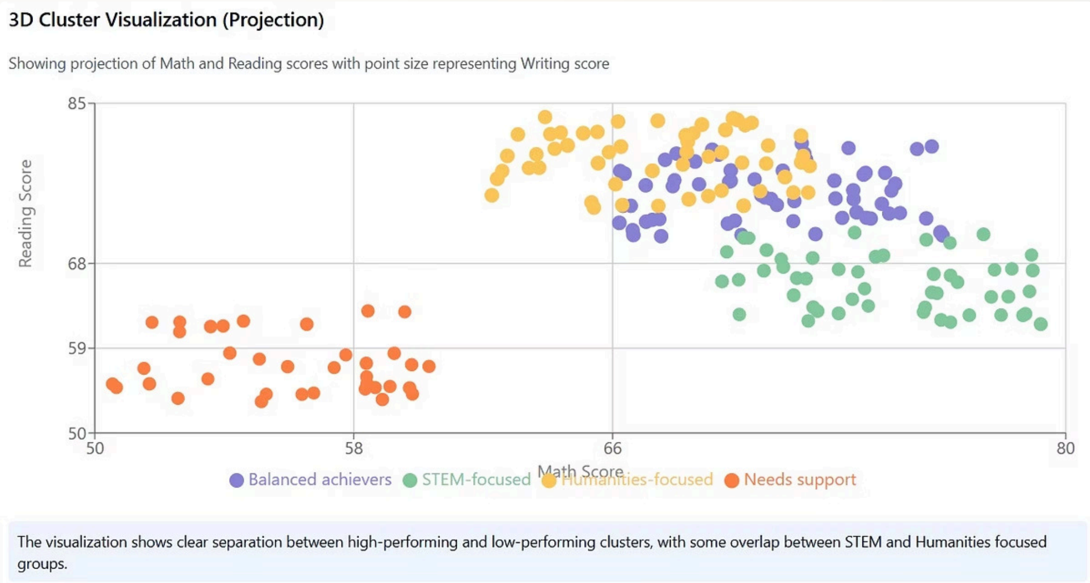
Next step: targeted interventions based on these insights.



Algorithm Performance Metrics Comparison

1

- Stacking ensemble shows best overall balance
- Gradient Boosting and XGBoost demonstrate strong precision and recall
- Random Forest has competitive accuracy but lower F1-Score
- Decision Tree performs adequately but is outperformed by ensemble methods



Student Performance Profile Clusters

1. **Explore student performance patterns with a 3D cluster visualization of Math and Reading scores.**

2. **Uncover four distinct groups:**
   - **Balanced achievers**: Consistent performance across subjects
   - **STEM-focused**: Higher math, lower reading scores
   - **Humanities-focused**: Higher reading, lower math scores
   - **Needs support**: Lower performance across all areas

3. **Discover insights:**
   - Clear separation between high and low-performing clusters
   - Some overlap between STEM and Humanities focused students

4. **Insight:**
   - Student performance patterns reveal natural learning affinities that could inform personalized teaching approaches

Integrated Analysis & Action Plan

1 **Key Integration Points:**

- Student clusters align with gender performance differences identified earlier
- Socioeconomic factors (lunch type) linked to preparation differences
- Parental education impacts align with cluster performance patterns
- Ensemble modeling provides best identification of at-risk students

2 **Recommended Actions:**

- Develop targeted interventions for each cluster
- Address test preparation gaps, especially for lower socioeconomic groups
- Create parent engagement programs for academic support
- Utilize machine learning to identify at-risk students early
- Implement personalized learning paths based on subject strengths