Plymouth High School Data Analysis & Insights

February 2025

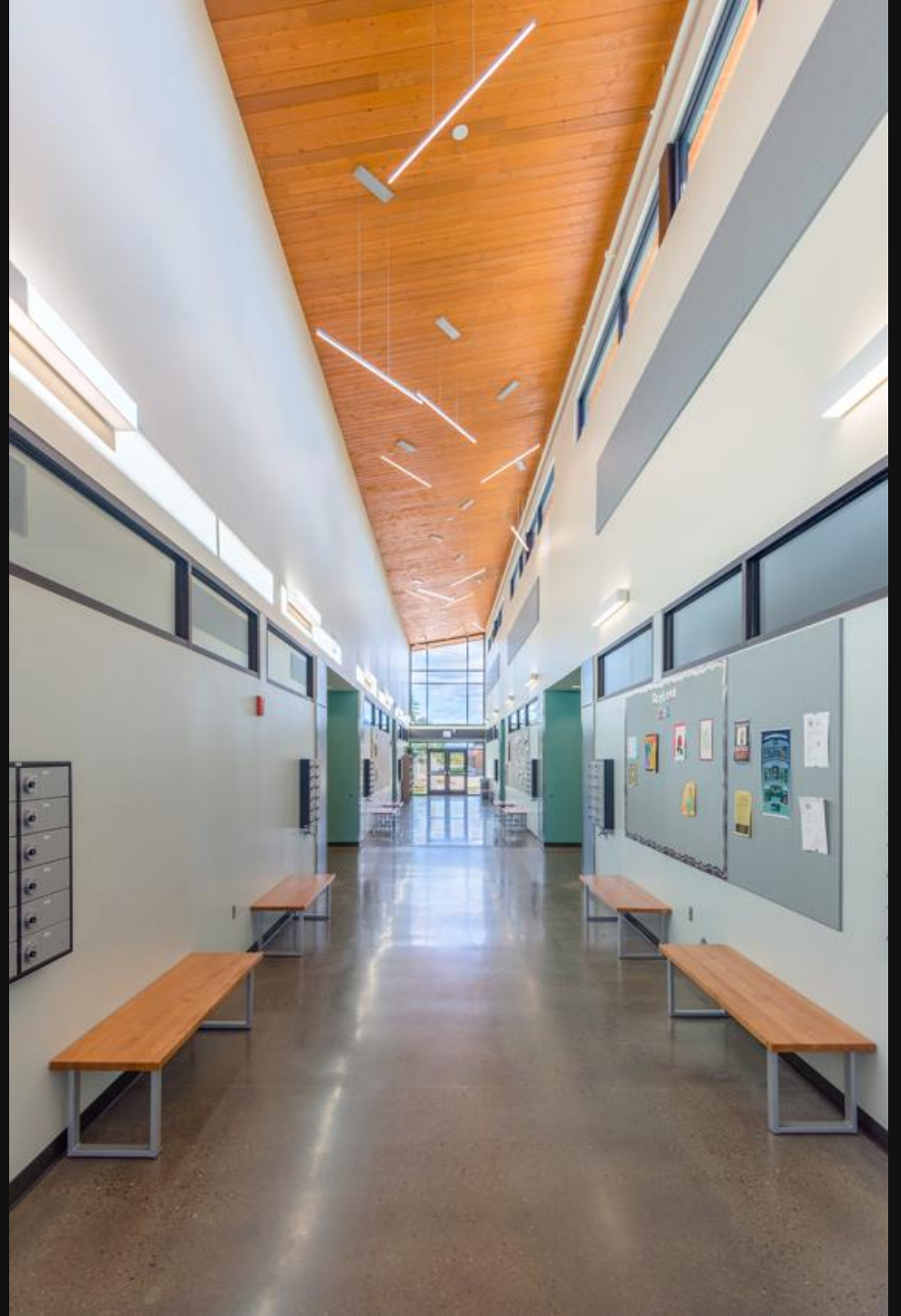


Table of Contents

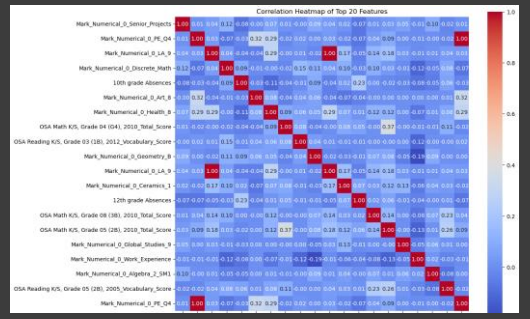
01.

Executive Summary

Deeply understanding key moments in students' high school experiences is key to powering a positive educational outcomes and administrative decisions.

Executive Summary

Deeply understanding key moments in students' high school experiences is key to powering a positive educational outcomes and administrative decisions.

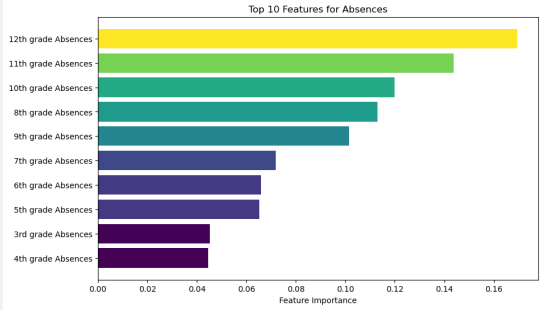


Why This Matters

This project aimed at using machine learning to improve two key areas – Regular Attenders and On-Time Graduation -- at Plymouth HS.

Synergy & Student Transcripts

Analyzing trends in student performance can led to targeted actions to improve overall graduation rates.



Improving Graduation Rates

Final Grades, Yearly Attendance, and State Testing provide insights into which factor has more importance for graduation.

Final Grades (7-12)

- The most impact we as educators can have on our students is their performance in the classroom.
- There are specific classes such as Senior Projects, PE 100, LA 10, and Discrete Math that had the most effect size for graduation.
- When Grades were analyzed separately, Citizenship, LA 12, Global Studies, and Geometry had sizeable impacts as well.

Yearly Attendance (K-12)

- However, if students are not present in the classroom, they're not able to make progress.
- When all factors were combined, absences during 10th grade had the most importance with 12th grade absences ranking as the second-most.
- However, when Absences were isolated and analyzed, the importance ranked 12th, 11th, 10th, 8th, and 9th, respectively.

State Testing (K-12)

- The combination of Final Grades and Attendance strongly influenced State Testing scores.
- Based on the model, State Testing was the category that had the most importance and variance, which makes it difficult to analyze.
- Two tests stood out: OSA Reading K/S Grade 3 Demonstrate Understanding and SBA Math HS Communicating Reasoning.

02.

Background and Context

The Student Transcript provides insights into positive and negative educational experiences throughout secondary level schools, which can be analyzed for commonalities and opportunities.

Trends ▶ Compared to Like-Schools in Oregon, Plymouth has comparable data in Regular Attenders and On-Time Graduation. Data for Like-Schools is inconsistent and limited.

Plymouth Comparison Averages 2013 - 2024

38% of students are **Regular Attenders,** which is students who attended more than 90% of their enrolled school days.

28% of students are **On-Time Graduates**, which is students earning a diploma within four years.

Source: Oregon At-A-Glance School and District Profiles and Accountability Details

Oregon & Like-School Comparison Averages 2013 - 2024

70%* of students across **Oregon** are considered **Regular Attenders**.

**[Like-School Data not available]*

27% of students in **Like-Schools** such as Alternative Schools are considered **On-Time Graduates**.

Source: Oregon At-A-Glance School and District Profiles and Accountability Details

The data available for Plymouth High School, all schools across Oregon, and Like-Schools in Oregon show that Plymouth is **consistent** with the state averages for student success.

Data Preparation, Caveats, and Project Methodology

This project uses a random forest classifier to explore how different factors like final course grades, yearly attendance, and state testing results, relate to high school graduation. The data was processed to ensure it was clean and ready for analysis, and the machine learning model was trained to predict whether a student is likely to graduate.

Data

Final Course Grades 7 – 8

Number of Middle School Credits Attempted

Number of Middle School Credits Completed

Final Course Grades 9 – 12

Number of High School Credits Attempted

Number of High School Credits Completed

State Testing K – 12

Yearly Attendance K – 12

Number of Students: 400

Caveats

These metrics are useful indicators of academic performance but do not capture the full complexity of student success.

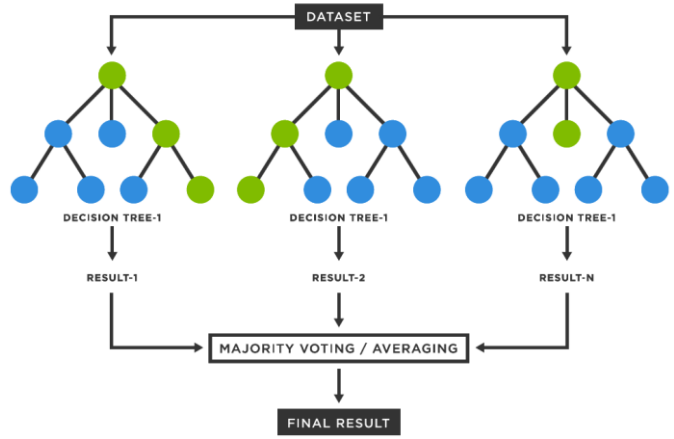
Academic records reflect **observable performance** but miss important contextual factors such as socioeconomic background.

A holistic view of a student includes **their aspirations, resilience, and unique challenges**, none of which are fully captured by attendance, grades, or test scores alone.

Graduation rates are affected by a complex interplay of **academic, social, and personal** dynamics that are not represented in the data.

Data for each student is broad but not all-encompassing. The lack of data due to a variety of factors means that some variables – like attendance or state testing – have some variance.

Methodology



03.

Results and Insights from Data

Connecting the dots and the data based on the model shows a clear theme of specific classes, number of days absent in certain years, and threshold scores on state tests that indicate whether a student will graduate from high school.

Overview of Results

The model predicted with **70% accuracy** whether a student graduated from high school.

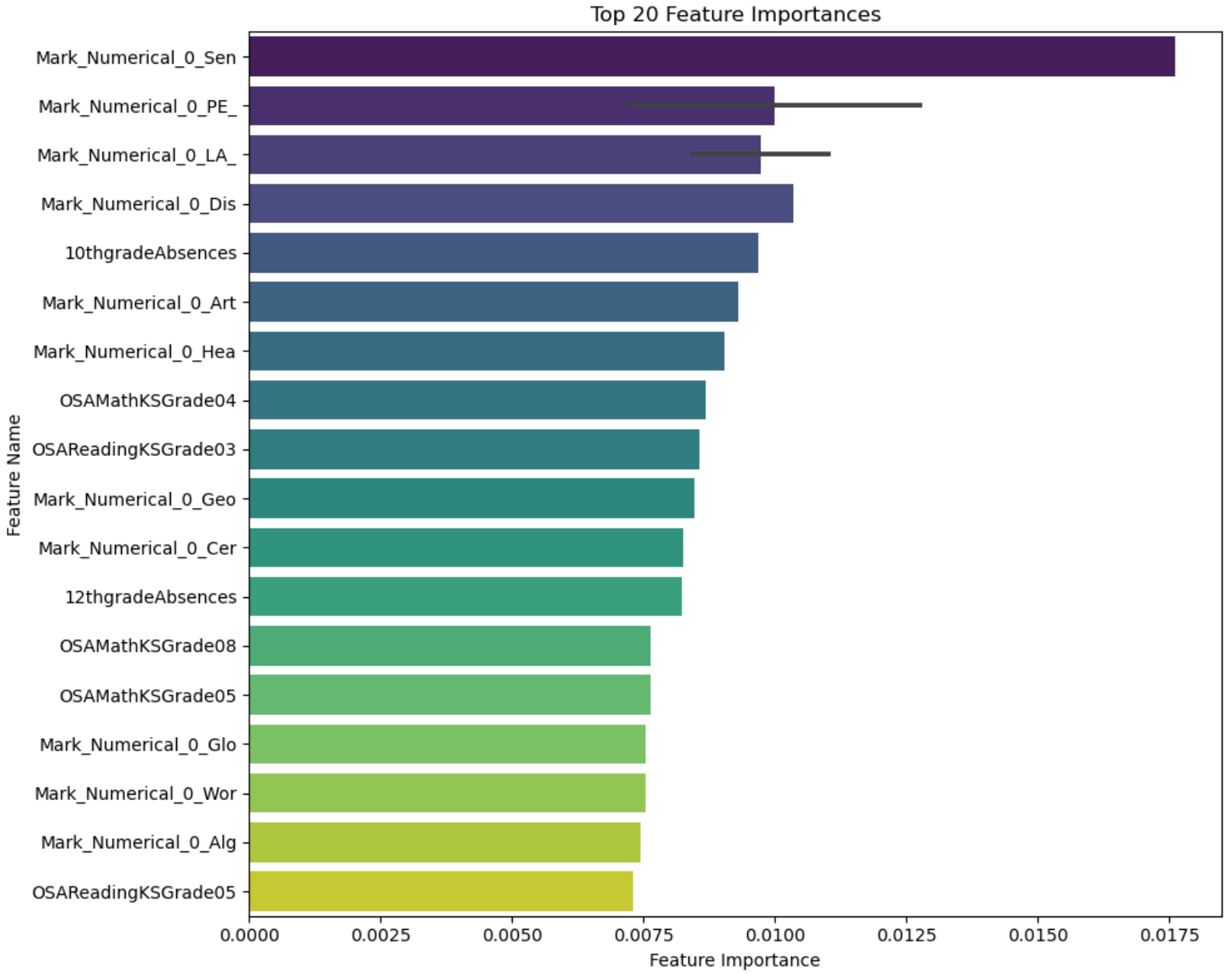
The model is better at predicting students who **won't graduate** than those who will.

Top Predictors of Graduation:

Senior Projects, PE 100, LA 10, and Discrete Math

Absences in 10th and 12th grades

Reading and Math assessments (Grades 3–5)

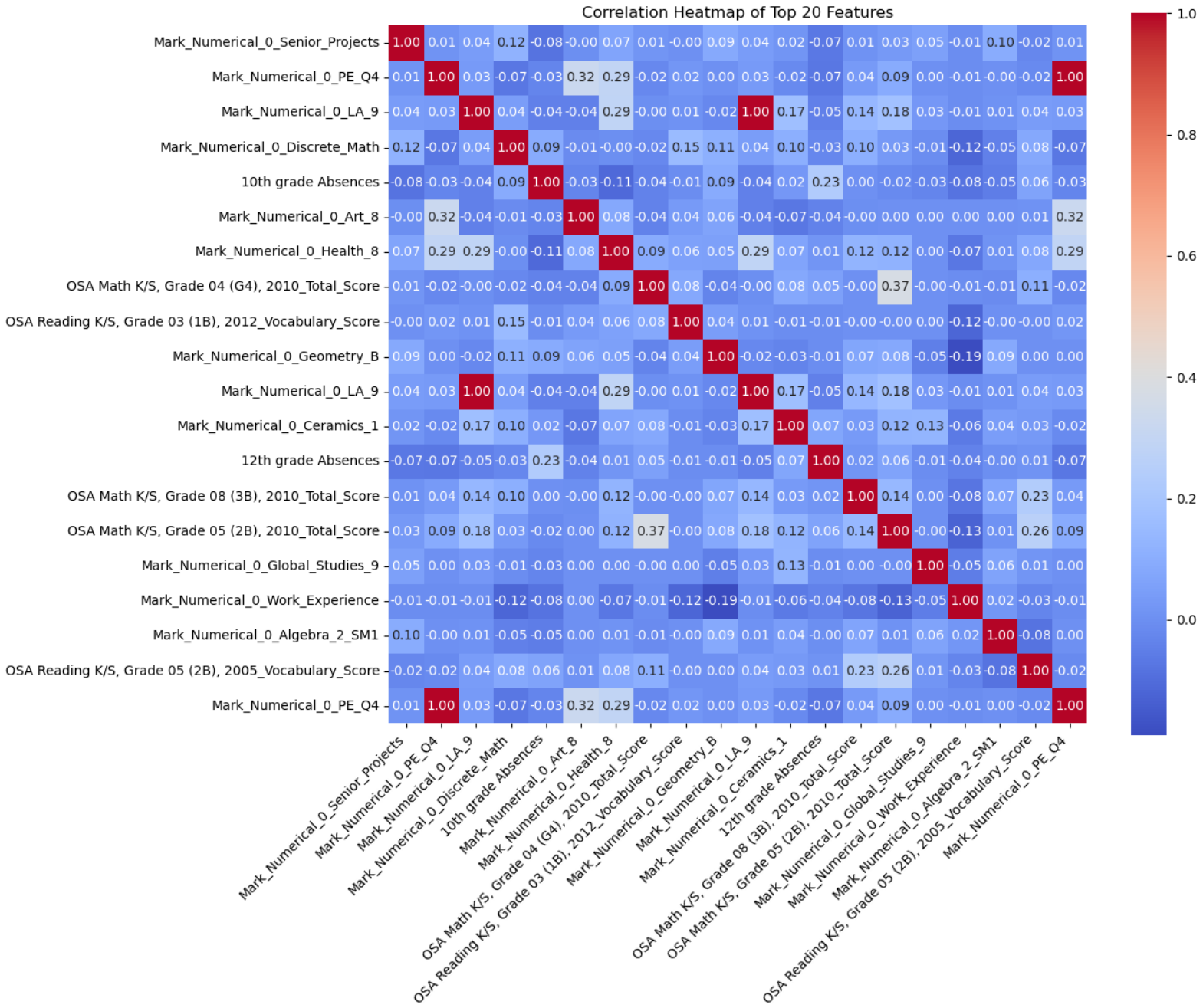


Overview of Results

In real-world data, features may be **highly correlated** with each other.

Feature importance measures the **predictive contribution** within the entire model, considering interactions and the joint effects of features.

In a broader context (when combining all features), consider how much a variable **contributes to reducing prediction error** or improving model performance when interacting with other variables.



Focus on Categories vs. Overall Results

Prioritizing Top 10 Features by Category Instead of the Overall Top 20

When analyzing the factors that impact high school graduation, we identified the top 20 features overall and also isolated the top 10 features within specific categories—final grades, yearly attendance, and state testing. While the top 20 list gives a big-picture overview, focusing on the top 10 features in each category is more actionable and meaningful for educators.

Category-Specific Insights Matter

By isolating features, we can understand what matters most within each area (grades, attendance, or testing).

This specificity helps us tailor interventions and strategies to address the root causes in each area rather than generalizing across them.

Clarity for Teachers and Administrators

Teachers are directly involved with grades, attendance, and test preparation. Seeing the most important features within their *area of influence* makes the data more relatable and actionable.

Actionability and Focus

While the top 20 list is comprehensive, it might be overwhelming or mix factors that educators have limited control over. Focusing on the top 10 features by category ensures that each list is both manageable and highly relevant to specific domains where teachers can make a difference.

Improved Collaboration Across Teams

When the data is broken down into categories, it encourages collaboration among staff members. Each group (e.g., grade-level teachers, administrators, etc.) can focus on their strengths while aligning efforts to support graduation collectively.

04.

Deeper Dive into Categories

Diving deeper into the categories and the data from the model unearths critical insights that will lead to takeaways and actions.

Terms

The classification report helps us measure how well our machine learning model makes predictions. In this case, the model is trying to predict whether a student will graduate from high school (True) or not (False).

Precision

Precision measures how often the model's predictions for a specific outcome are correct.

- For **False (not graduating)**: If the model predicts "not graduating," it's correct **70% of the time**.
- For **True (graduating)**: If the model predicts "graduating," it's correct **64% of the time**.

Think of it like aiming at a target—precision tells us how many of our hits were on the right target.

Recall

Recall shows how good the model is at catching all the students in each group.

- For **False**: The model correctly identified **91%** of the students who actually won't graduate.
- For **True**: The model only caught **28%** of the students who will graduate.

It's like a safety net—how much of the group we actually captured.

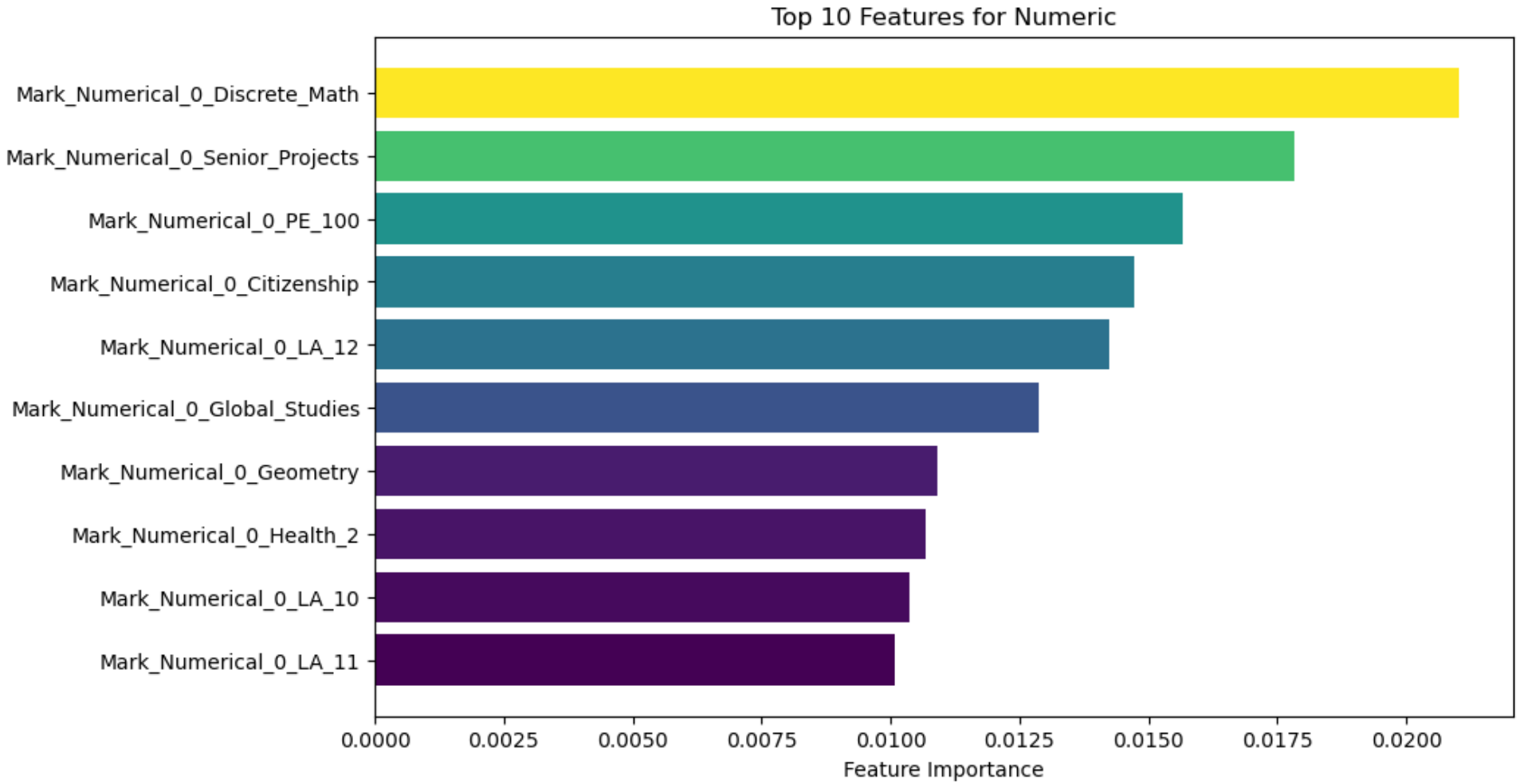
Final Course Grades (7 – 12) & Graduation Status

When only using final course grades to predict graduation status, the model had **70% Accuracy** overall for predicting graduation status with **92% Recall** ensuring most at-risk students were correctly flagged.

Key Insights

When Final Course Grades were analyzed separately, the following courses had sizeable impact as well:

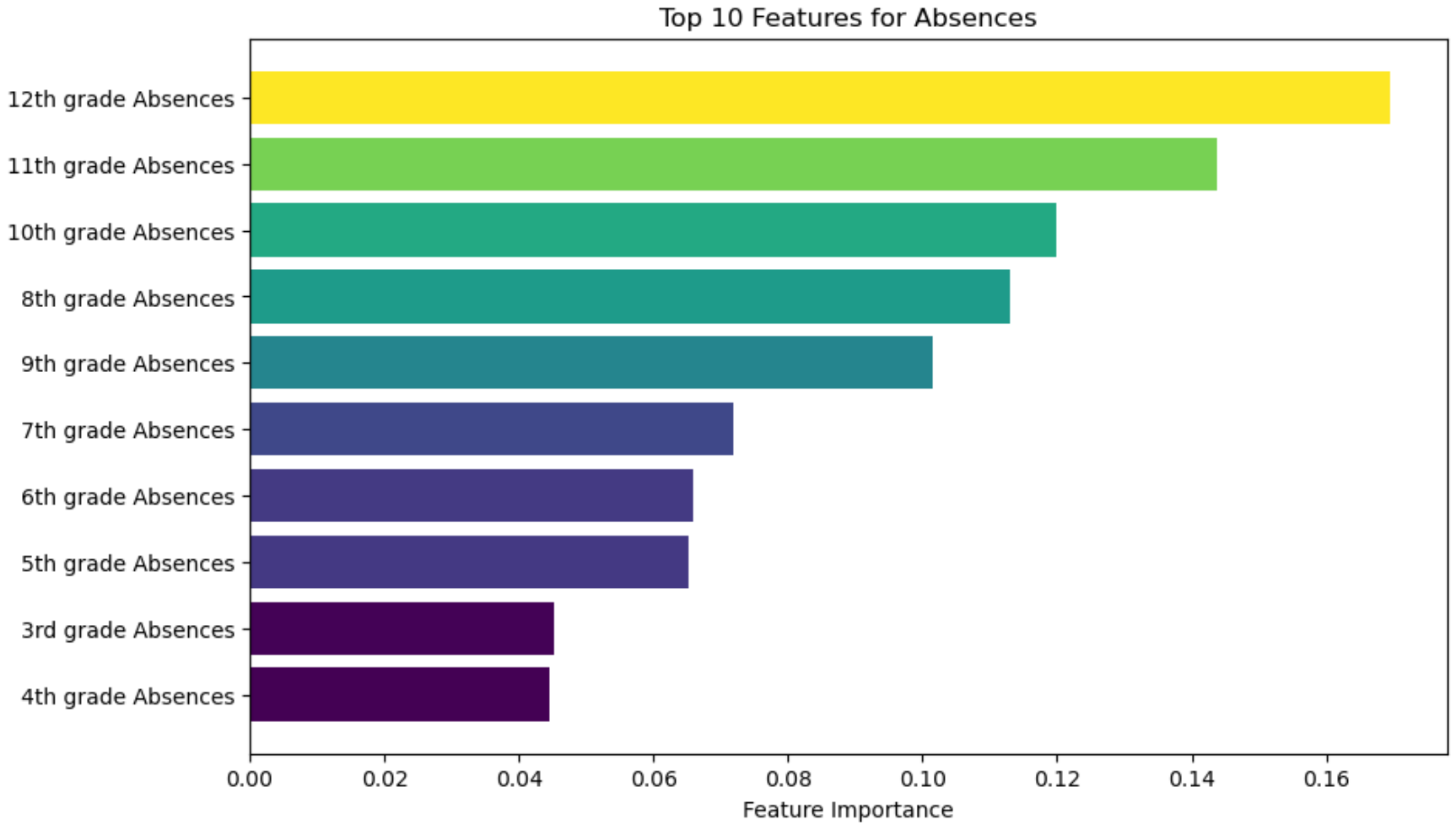
- Citizenship
- LA 12
- Global Studies
- Geometry



Yearly Absences (K - 12) & Graduation Status

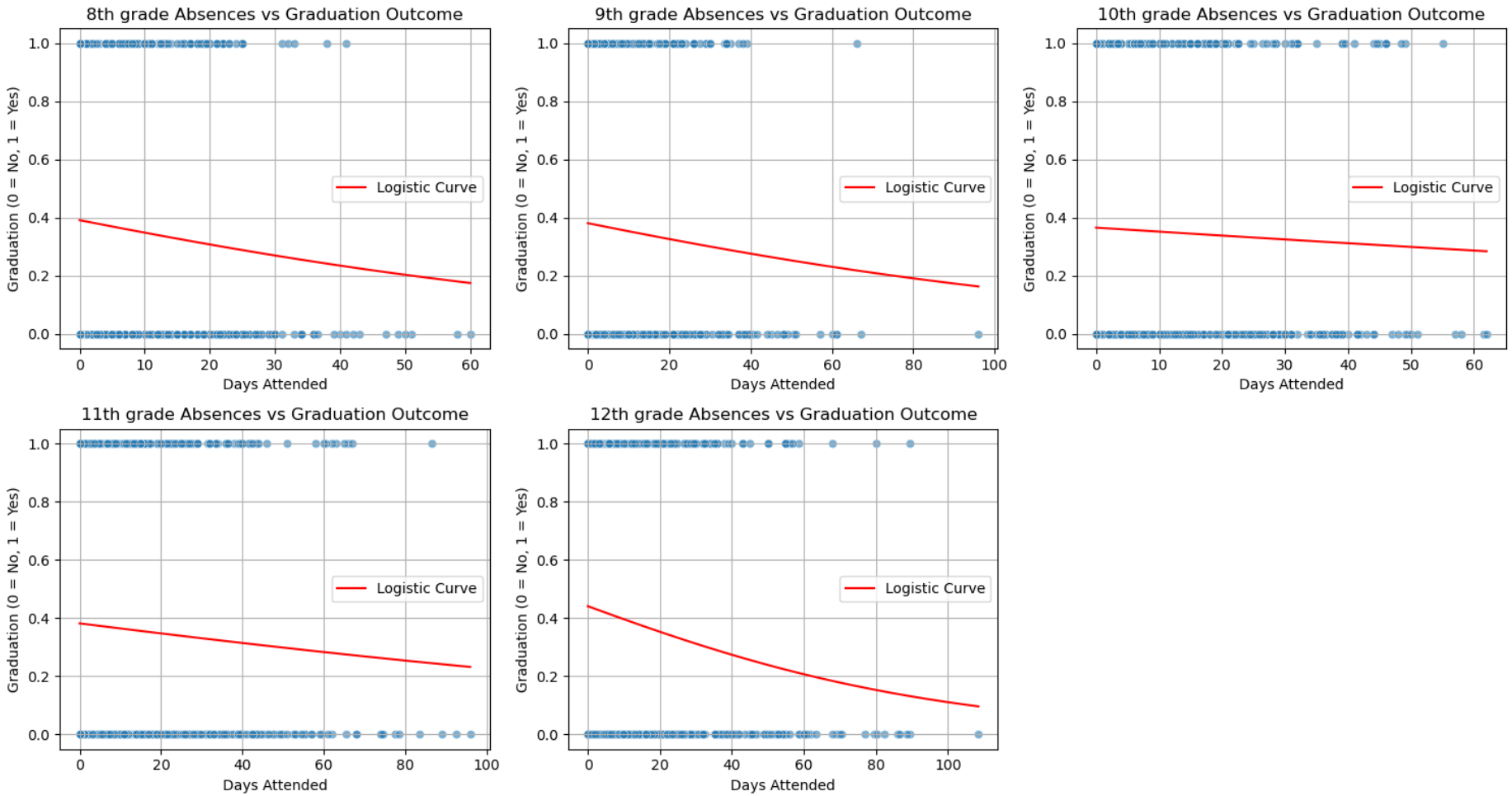
The model is more reliable at flagging at-risk students based on their absences**. 68% (Precision)** of the flagged at-risk students are truly at risk while **90% (Recall)** of at-risk students are successfully identified.

Full data from kindergarten through 12th grade is **limited**, though there seems be **less correlation** between elementary and middle school attendance and high school graduation.



Top 10 Features for Absences (K – 12)

However, its ability to predict likely graduates is limited (**Precision: 56%**, so only just over half of predicted graduates are correctly identified), potentially due the relationship between absences and graduation.



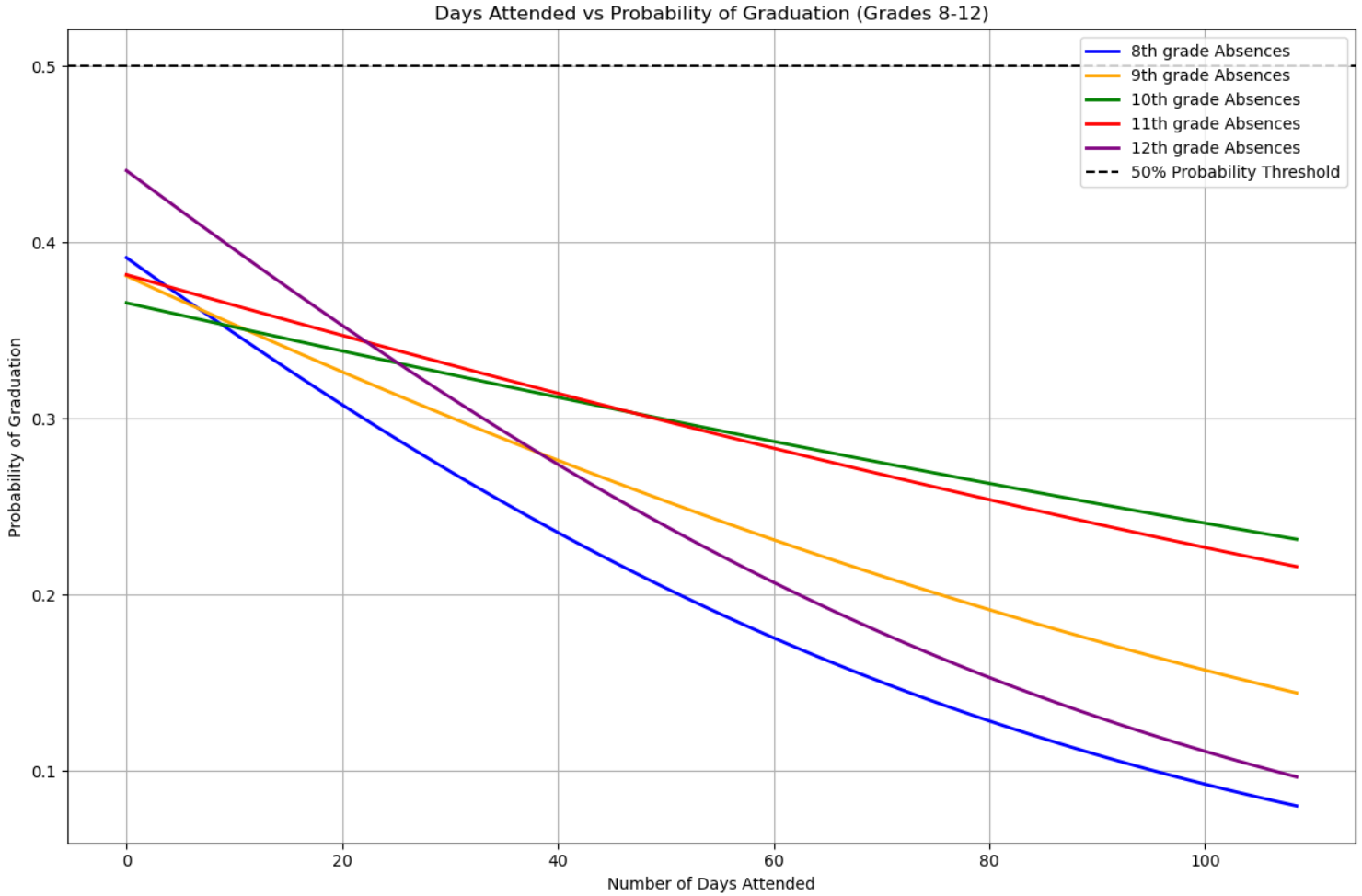
Top 10 Features for Absences (K – 12)

When trying to visualize absences and probability of graduation, the model struggles to create specific **thresholds** for the **minimum number** of days attended each year for the student to most likely graduate.

Key Insights

The y-intercept for 10th grade absences is the **lowest**, indicating sizeable impact.

For 11th and 12th grade, the model found that after **15 absences**, the probability of graduation for that student starts to **drop below 50%.**



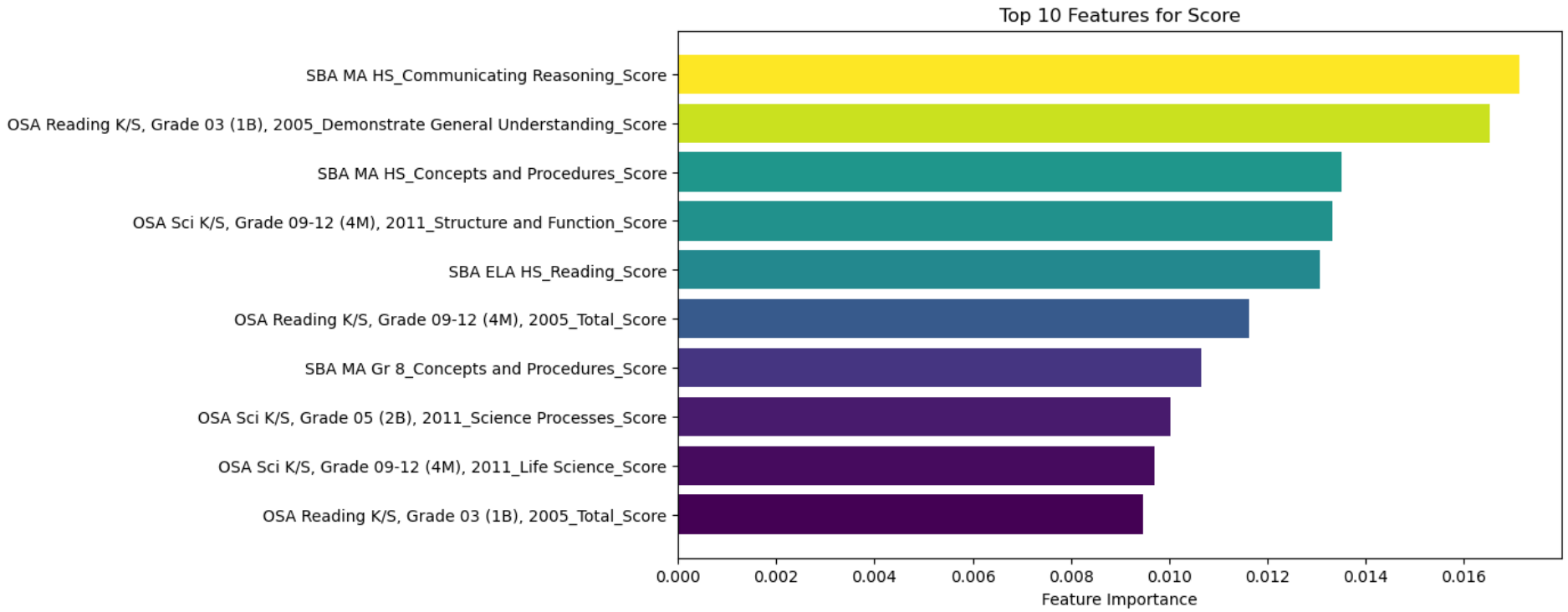
Top 10 Features for State Testing (K – 12)

While State Testing was one of the most important features, it has the lowest **Accuracy (63%)** meaning that even though it has influence on the model's decision-making, it's not making consistent predictions.

Key Insights

While **SBA High School Math** was not among the combined top 20 features, its tests have significant weight when analyzed independent.

The OSA Grade 3 Reading score had an impact **across all features**, the most being the Discrete Math grade



State Testing Confusion Matrix

A confusion matrix is a table used to evaluate the performance of a classification model

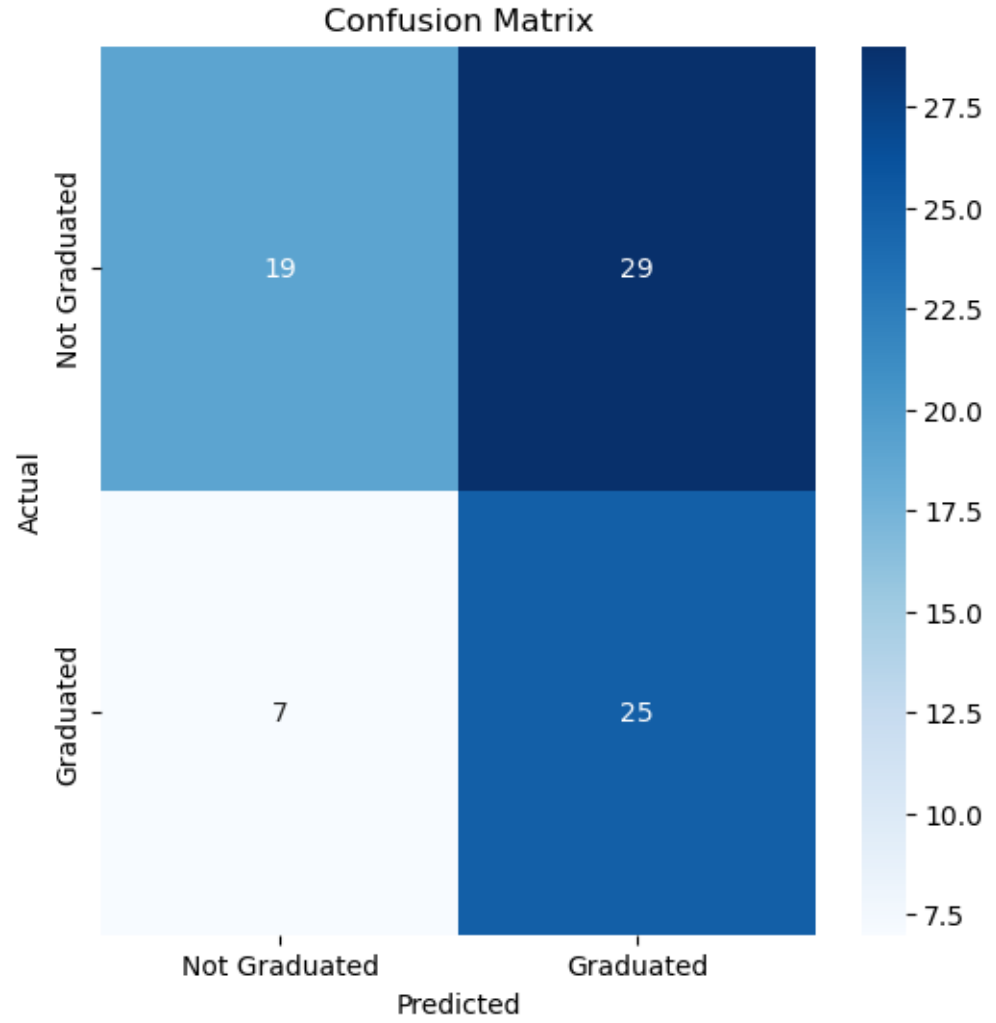
Positives and Negatives

Top-left: True Negative (TN) – Correctly predicted negative cases

Top-right: False Positive (FP) – Incorrectly predicted as positive

Bottom-right: True Positive (TP) – Correctly predicted positive cases.

Bottom-left: False Negative (FN) – Incorrectly predicted as negative)



Only 56% of students were correctly grouped into either True Negative or True Positive.

The model struggled to predict graduation solely based on State Testing scores, incorrectly predicting **36% False Positives**.

Scores closest to a 50% probability of graduating:

OSA Reading K/S, Grade 03

202

Note: 204 is general Meets Performance Level Benchmark

SBA MA HS

2480

Note: 2540 is generally Level 2 out of 4 for Level Benchmark

05.

Takeaways and Future Research

The insights point towards three primary areas for focus and impact, as well as and further intervention opportunities.

Main Areas for Focus and Impact

1

Rethink, Not Reallocate

There are evidenced-based interventions in-place for a variety of students in overlapping demographics. Changing this setup would be detrimental to the foundation the staff has already created. Instead, rethink efforts towards students who do not specifically have targeted interventions, like 10th graders.

2

Data-Based Prioritization

While students have strengths and weaknesses, if the staff needs to decide for a student to prioritize one class over another, take into account the impact of each class in the data while examining the student's course history to set up the best chance of success for Credit Recovery.

3

Worry Less, But Plan More

There are only some many variables the staff can control with regards to students graduating high school. The students come to you near the end of their education career and the best efforts are put forth to make sure that they're successful.

However, if there are initiatives to help support K – 8 students through mentorship, then the high school students can have a larger impact earlier in educational timelines.

THANK YOU

For more information, questions, or future project proposals, please contact mitchk@sthelens.k12.or.us