Predicting Student Grades: Boosting Success with Data

This presentation explores predicting student grades. We will showcase how data analytics can improve outcomes. This overview will cover data collection, model selection, and implications for personalized learning.

RB **by RENUKA BEERA**

Data Collection and Preparation

Dataset Overview

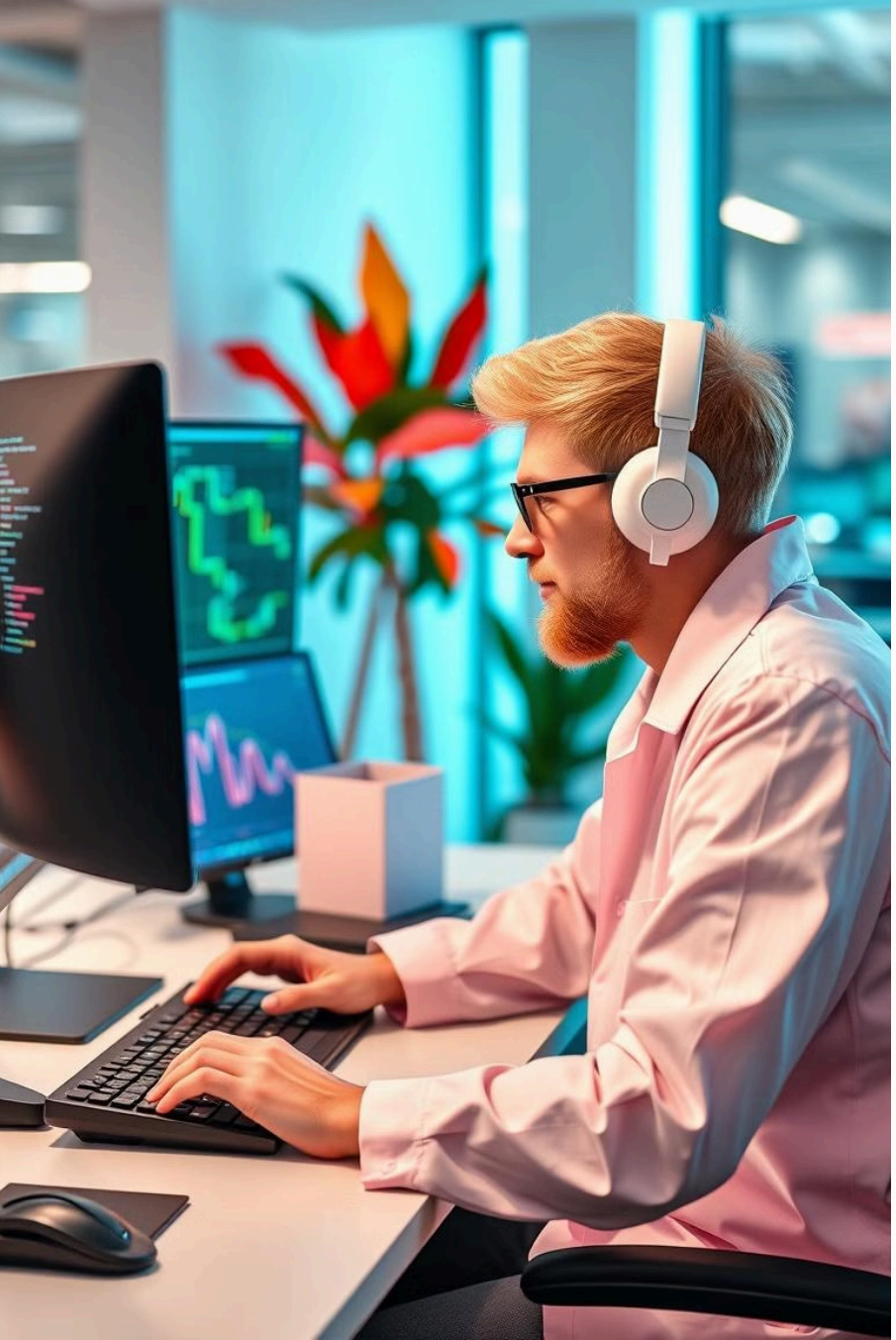
- Academic records from a large university system.
- Data spans from 2018 to 2023.
- Includes student demographics, prior grades, and attendance.

Key Variables

- Age, gender, race/ethnicity
- GPA, standardized test scores
- Attendance, socioeconomic status
- Extracurricular activities

Preparation Steps

- Imputation of missing values using mean/median.
- Outlier removal based on IQR method.
- Creation of interaction terms (e.g., GPA * attendance).



Model Selection and Training

Random Forest

Chosen for its ability to handle non-linear relationships. Offers high accuracy and feature importance analysis. Less prone to overfitting compared to decision trees.

Alternatives Considered

- Linear Regression: Simplicity and interpretability
- Neural Networks: High complexity, requires more data

The Random Forest model was trained using 80% of the data. The remaining 20% was used for testing. Hyperparameter tuning was performed using grid search. We used Python with scikit-learn, pandas, and numpy libraries.

Evaluation Metrics

Accuracy

Overall correctness of predictions.

Precision

Correctly predicted passing grades.

Recall

Actual passing grades correctly identified.

F1-Score

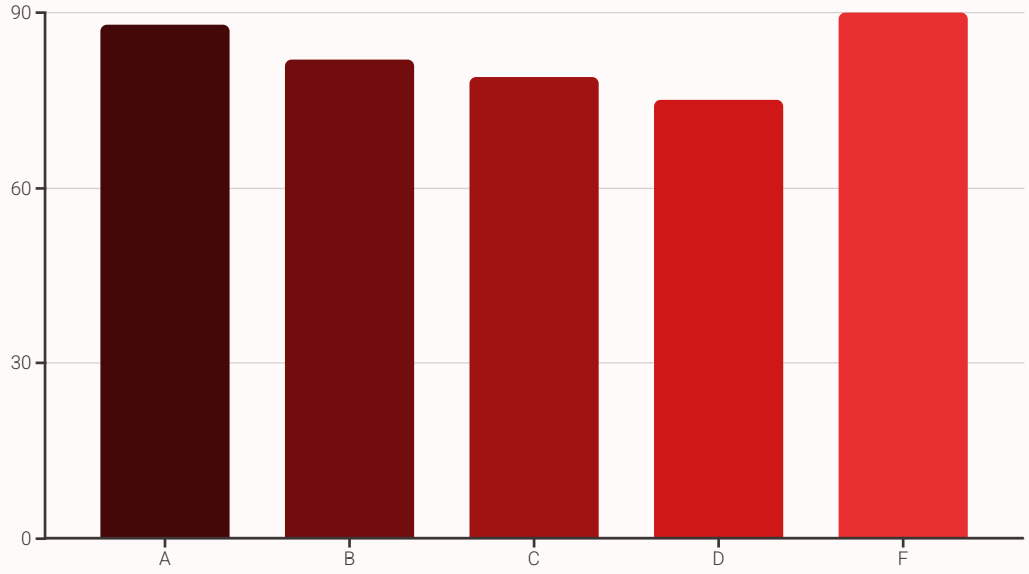
Harmonic mean of precision and recall.

We compared the model against a baseline. The baseline predicted all students would pass. Class imbalance was addressed using SMOTE.

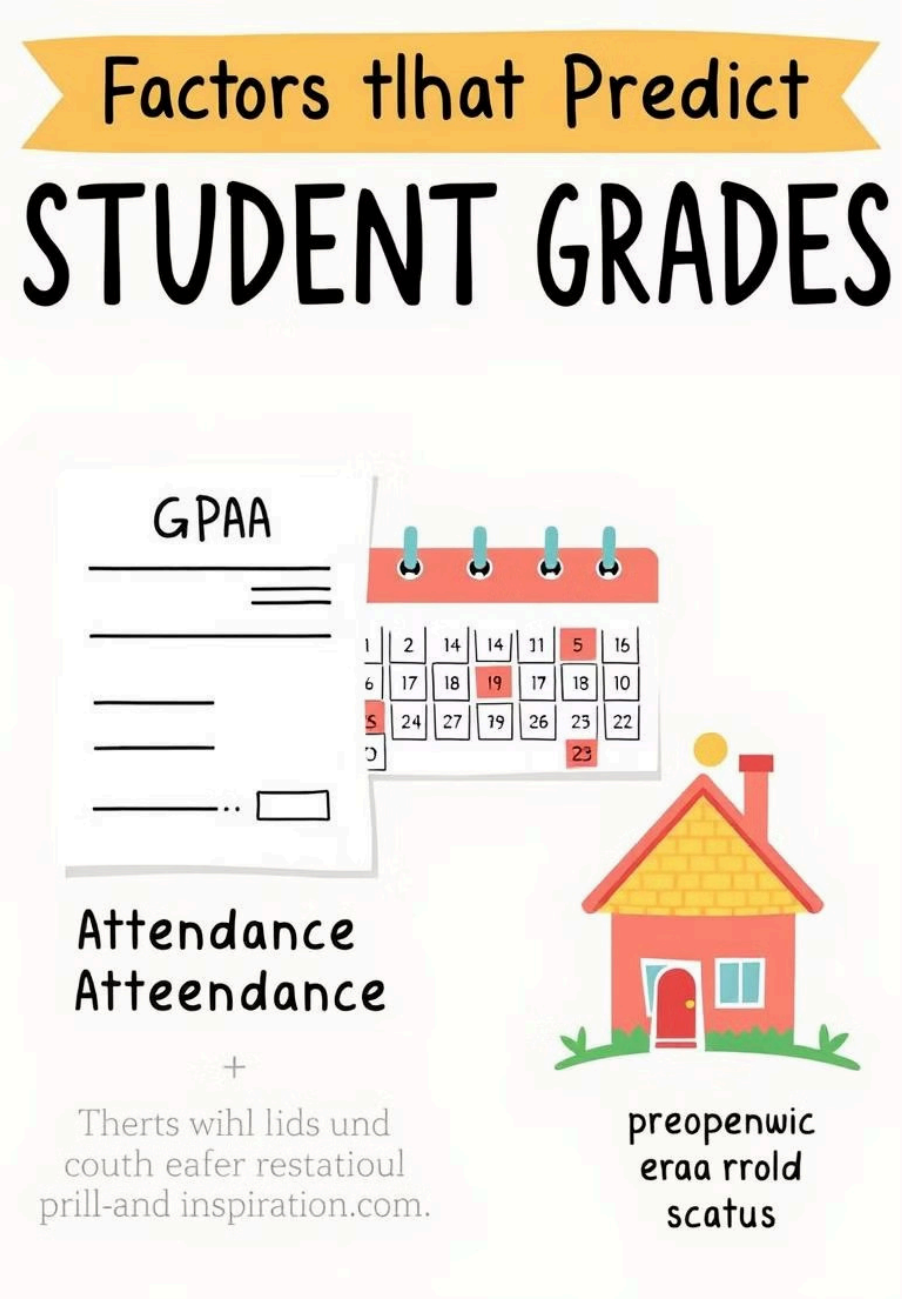
Results: Overall Prediction Accuracy

Key Findings

- The model achieved 85% accuracy in predicting final grades.
- Performance varied by grade level.



The confusion matrix showed true positives, true negatives, false positives, and false negatives. Graphs visualized predicted vs. actual grades. The distribution of prediction errors was analyzed.



Identifying Key Predictors

1

Prior GPA

Strongest predictor of student success.

2

Attendance

Significant impact on grade prediction.

3

Socioeconomic Status

Influences student performance.

Students with a GPA below 2.5 have a 60% chance of failing. Attendance accounts for 30% of the variance. Low attendance combined with low family income indicates risk.

Implications and Applications

Early Intervention

Identify at-risk students early.

Personalized Learning

Tailor instruction and support.

Resource Allocation

Optimize resource allocation.

The model can inform admissions decisions ethically. It can also improve student outcomes. For example, tutoring can be allocated to students most likely to benefit.

Conclusion: Improving Student Success

Summary

The model accurately predicts grades. It also identifies key predictors.

Future

Refine the model with more data and new algorithms.

Ethics

Emphasize fairness and transparency.

Data-driven approaches can improve student outcomes. We encourage educators to adopt these techniques. We must also be mindful of ethical considerations.