STUDENTS' PERFORMANCE ANALYSIS

Mattias Minejev
Annabel Aleksius
Tõnis Tõnissoo

Goal 1: Train a model to predict how will the student perform academically

Goal 2: Analyse what kind of students perform the best

Goal 3: Analyse which factors impact which test results the most.

Data science methods

- Random forest classifier
- Linear SVM
- KNN
- Logistic regression
- XGBoost
- Correlation matrix
- Scaling
- Aggregation
- Comparison
- Visualization

Kaggle's Dataset (56.62 kB):

public dataset, containing information about students' demographic, parental education and academic performance information.

📚 Students Performance Dataset 📚

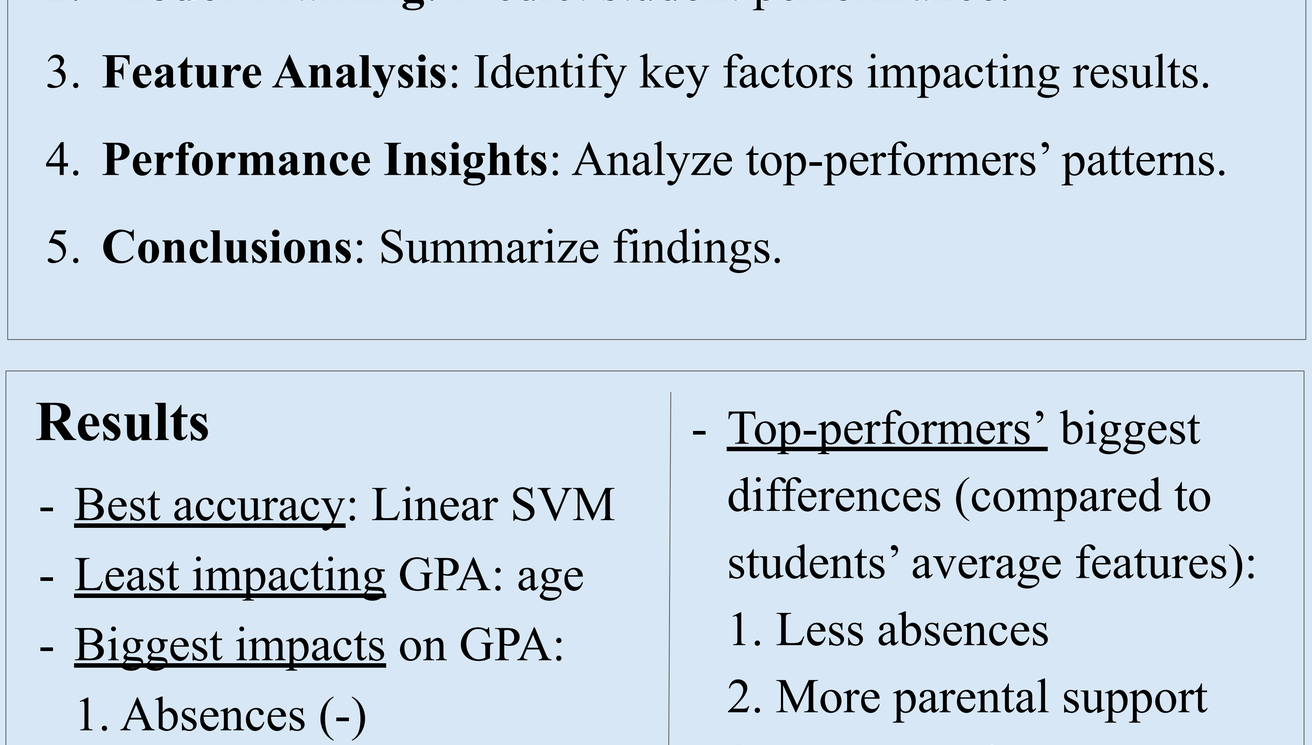
Approach

1. **Data Preparation**: Clean the dataset and split the data into training and test sets.
2. **Model Training**: Predict student performance.
3. **Feature Analysis**: Identify key factors impacting results.
4. **Performance Insights**: Analyze top-performers' patterns.
5. **Conclusions**: Summarize findings.

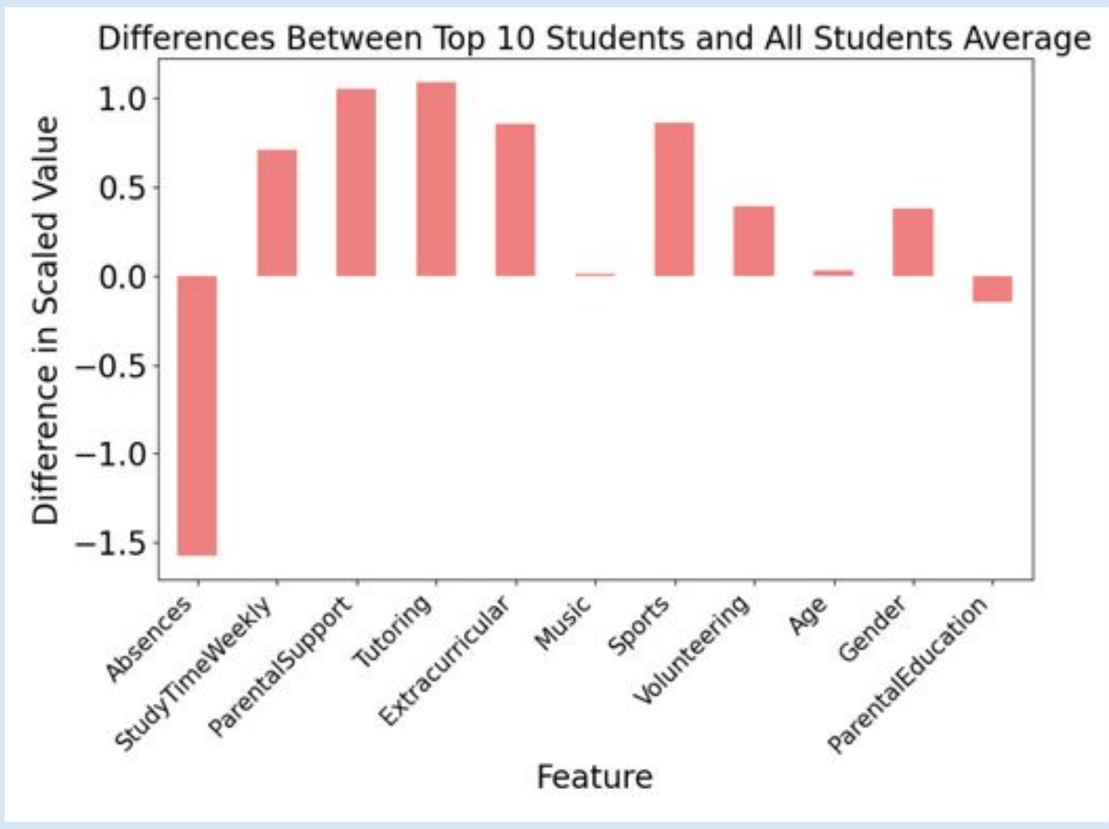
Results

- Best accuracy: Linear SVM
- Least impacting GPA: age
- Biggest impacts on GPA:
  1. Absences (-)
  2. Parental support (+)
  3. Weekly study time (+)

- Top-performers' biggest differences (compared to students' average features):
  1. Less absences
  2. More parental support
  3. More tutoring
  4. More extracurriculars
  5. More weekly study time



Top-performers' differences



Correlation