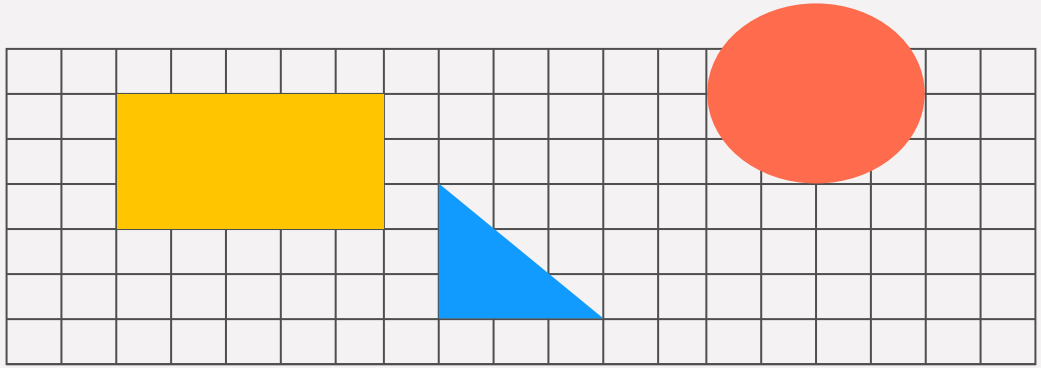
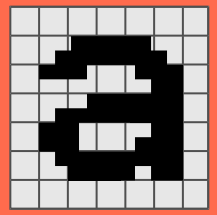
Akash Pamal, Maseel Shah, and Charlotte Zhang

Machine Learning For Human Learning

Using machine learning algorithms to determine the best measure of student success in Virginia

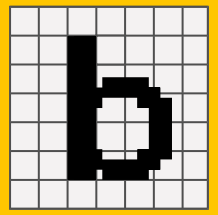


Driving purpose



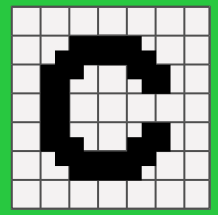
Machine learning

How can we leverage machine learning to predict success in Virginia high schools?



Best factors

Which factors are the most correlated with academic success?



Success metric

Which metric is the best for predicting academic success: graduation rate or standardized test scores?

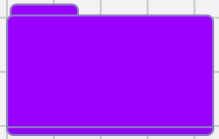
The data

Demographics

Economic Factors

Student Behaviors

Teachers



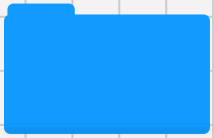
Testing

Virginia public schools

- Kaggle – Virginia Department of Education
- School year 2021-2022
- 19 files
- 119 columns

Data cleaning

Demographics



Economic Factors

Student Behaviors

Teachers

Testing

Virginia public schools

- Primary key (sch_div)
- Pivoted data
- 285 high schools
- 35 features

Pivoting data

Demographics/Race.csv

Race	Percentage	Sch_Div
Black, not of Hispanic origin	44.60784314	Accawmacke Elementary Accomack County Public Schools
White, not of Hispanic origin	27.94117647	Accawmacke Elementary Accomack County Public Schools
Hispanic	24.26470588	Accawmacke Elementary Accomack County Public Schools
Non-Hispanic, two or more races	1.715686275	Accawmacke Elementary Accomack County Public Schools
Asian	1.470588235	Accawmacke Elementary Accomack County Public Schools
Black, not of Hispanic origin	36.97234352	Arcadia High Accomack County Public Schools
Hispanic	31.58660844	Arcadia High Accomack County Public Schools
White, not of Hispanic origin	29.54876274	Arcadia High Accomack County Public Schools
Asian	1.164483261	Arcadia High Accomack County Public Schools
Non-Hispanic, two or more races	0.727802038	Arcadia High Accomack County Public Schools
Black, not of Hispanic origin	36.38025594	Arcadia Middle Accomack County Public Schools
Hispanic	32.17550274	Arcadia Middle Accomack County Public Schools
White, not of Hispanic origin	28.51919561	Arcadia Middle Accomack County Public Schools
Non-Hispanic, two or more races	2.376599634	Arcadia Middle Accomack County Public Schools
Asian	0.365630713	Arcadia Middle Accomack County Public Schools
American Indian or Alaska Native	0.182815356	Arcadia Middle Accomack County Public Schools

Sch_Div	Black, not of Hispanic origin	Hispanic	White, not of Hispanic origin	Asian	Non-Hispanic, two or more races
Arcadia High Accomack County Public Schools	36.97234352	31.58661	29.54876274	1.16448	0.727802038

Models

Linear Regression

Used as a baseline model to determine correlation between features to predict labels

Random Forest

More complex model to help make predictions in the event of non-linear and more complex relationships

Support Vector Regression

Used to identify complex and non-linear relationships, using the kernel trick and handling outliers

Graduation Rates

Linear regression

Root mean squared error

Train: 3.88 Test: 3.11

Random forest

Root mean squared error

Train: 4.31 Test: 3.05

Optimization

Grid search CV

Best parameters

max_depth: 10
min_samples_leaf: 1
min_samples_split: 2
n_estimators: 300

Support vector regression

Root mean squared error

Train: 3.92 Test: 3.07

Optimization

Grid search CV

Best parameters

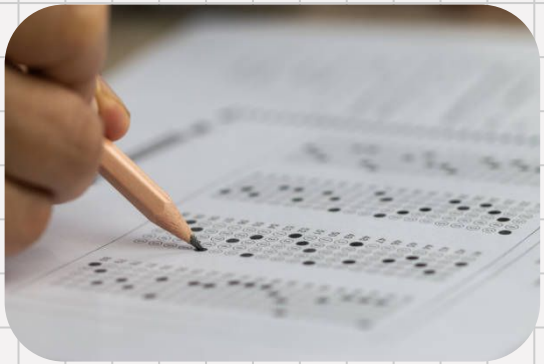
C: 100
epsilon: 0.01
gamma: 0.01
kernel: rbf

Standardized Test Scores

Linear regression

Root mean squared error

Train: 6.78 Test: 6.96



Random forest

Root mean squared error

Train: 7.68 Test: 6.67

Optimization

Grid search CV

Best parameters

max_depth: None,
min_samples_leaf: 4
min_samples_split: 2
n_estimators: 100

Support vector regression

Root mean squared error

Train: 7.39 Test: 6.64

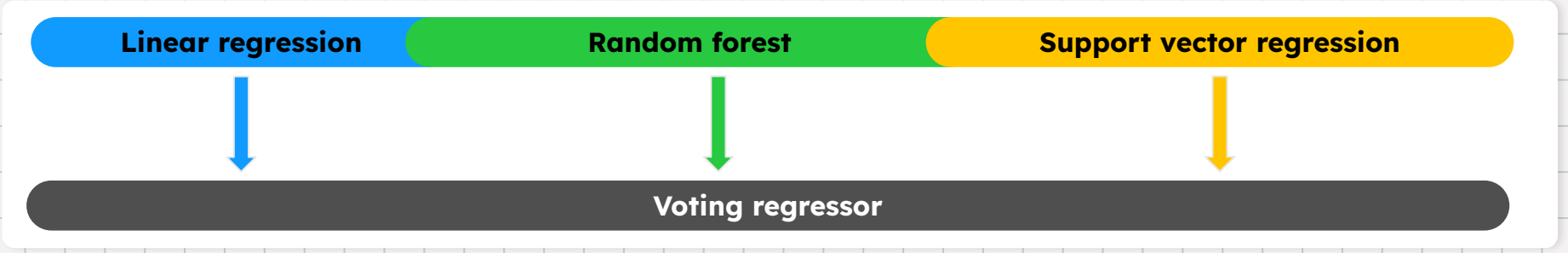
Optimization

Grid search CV

Best parameters

C: 100
epsilon: 0.01
gamma: 0.01
kernel: rbf

Ensemble Learning

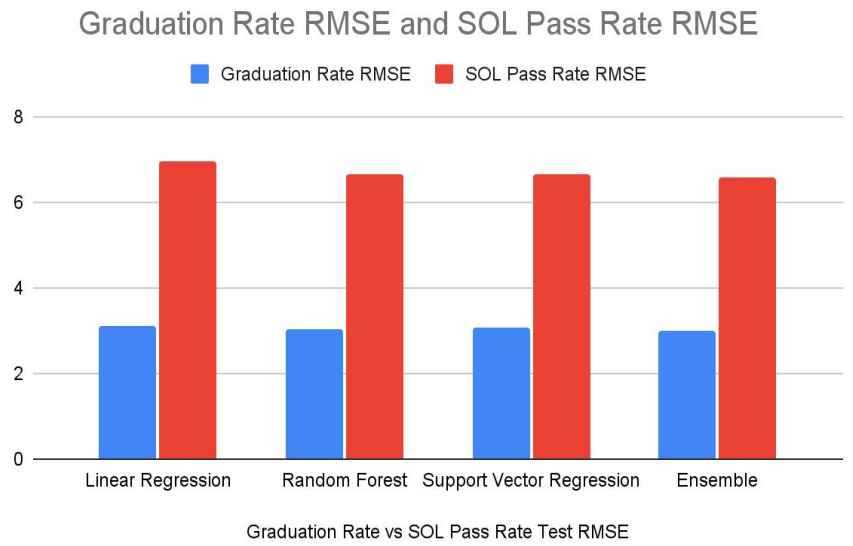


Root mean squared error

Test: 2.98 (graduation rate)
Test: 6.57 (SOL score)

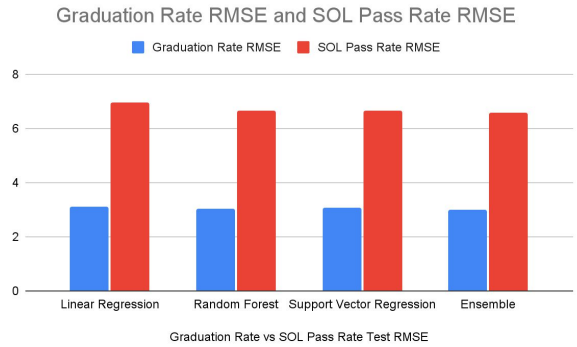
Results

Best model



Results

Best model



Most important features across models

Teachers_Bachelors_Degree_Percentage

Race_Black_Not_Of_Hispanic_Origin_Percentage

Homeless_No_Total_Count

Foster_Care_No_Total_Count

Disabled_No_Total_Count

Findings

- Our models performed better on graduation rate
- The most important features should guide how we improve our school systems
  - Support teachers pursuing higher education
  - Provide more resources to underprivileged groups
- Help provide students with targeted resources to be successful

What's next?

Encourage more studies like ours

- Consider more experimental studies rather than correlation

Translate results to real change in school systems

References

Data

https://www.kaggle.com/datasets/zsetash/virginia-public-schools?resource=download

Images

https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.istockphoto.com%2Fphotos%2Fhigh-school-graduation&psig=AOvVaw1FeCwof0Bu6mXezgrg6xKG&ust=1731956589183000&source=images&cd=vfe&opi=89978449&ved=0CBEQjRxqFwoTCLD18JuH5IkDFQAAAAAdAAAAABAE

https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.istockphoto.com%2Fphotos%2Fstandardized-testing&psig=AOvVaw0YEyAN_KAcUOlHVJ_TqdrT&ust=1731956860918000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCJD2-p-I5IkDFQAAAAAdAAAAABAE

https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.virginia.gov%2Fagencies%2Fdepartment-of-education%2F&psig=AOvVaw2Z-z704YBedYWNOxwKMVCh&ust=1732043495964000&source=images&cd=vfe&opi=89978449&ved=0CBcQjhxqFwoTCPC6-fzK5okDFQAAAAAdAAAAABAE

This work has been a part of the “Machine Learning for Virginia” project at the University of Virginia.