



# Final Project

# Sako Kassabian

CAN WE GET AHEAD OF “IN DANGER” STUDENTS?

# Education has an information problem

- ▶ Schools have a ton of data
- ▶ They have no idea what to do with it
  - ▶ It's actually a lot of industries
- ▶ It's all done after the fact with reports
  - ▶ Aggregate level
- ▶ Data must lead to information that leads to action

# Example of after the fact reports

## ► Caschooldashboard.com



# Data maturity in an organization

- ▶ Can't eat steak as a baby
- ▶ Executive leadership
  - ▶ We need analytics
  - ▶ We need machine learning
- ▶ Rise of black box tech companies

# Where does the data maturity journey begin?

- ▶ Consulting
  - ▶ Education for the educators
  - ▶ See where they are and tailor a solution
- ▶ Data cannot do their job for them
  - ▶ It is a tool, not a master
  - ▶ Not a silver bullet
- ▶ Limitations

# Getting in front of “In Danger” students

- ▶ With good data, no child slips through the cracks
- ▶ Focus your already stretched personnel
- ▶ Delve into the depths of aggregate reports

# Predicting “In Danger” students

- ▶ What is In Danger?
  - ▶ Converted a grade score to binary feature
    - ▶ If student English grade = D or F
      - ▶ Field = 1, else = 0

# Full Dataset

- ▶ Grade level
- ▶ E1, E2, E3
  - ▶ English standardized test scores for last 3 years
- ▶ Absent rate
- ▶ ELA (English Learning Acquisition) Status
- ▶ Race



# Feature Engineering

- ▶ English Grades conversion
  - ▶ A – F
    - ▶ -/+ grouped together
    - ▶ Into 1-5, else 0
- ▶ ELA (English Learning Acquisition) Status
  - ▶ One hot encoded via `pd.get_dummies`
  - ▶ EO, EL, IFEP, RFEP
- ▶ Race
  - ▶ One hot encoded via `pd.get_dummies`

# Feature Engineering Cont.

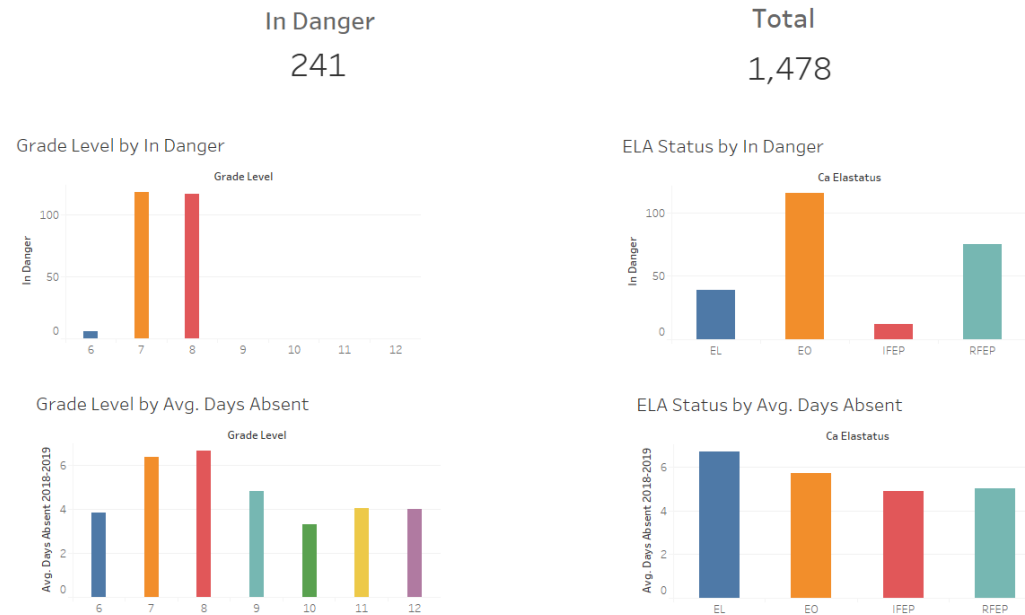
- ▶ Square
- ▶ Log
- ▶ Square root

# Models

- ▶ Logistic Regression
- ▶ Random Forest
- ▶ SVM
  - ▶ Models are ~90% accurate
  - ▶ The journey was tough – needed to overcome overfitting issues
  - ▶ Still struggling with specificity – more data required
- ▶ Mixed-effect model
  - ▶ Intraclass correlation of .78

# Final Output

- ▶ A dashboard that has a portion of your In Danger students
- ▶ Along with segmented data information where you have the ability to drill down and view those students



# Final Product – Models in Production

- ▶ Azure? Report?
- ▶ Delivery structure