Variable Star Classification

Phil Lannoo &
Connor McClain

Summary

- Business Problem
- Data Understanding
- Results
- Next Steps

Business Problem

- NASA wants to observe variable stars
- A star being variable could indicate interesting things
- Some stars have unclassified variability
- We've been tasked to predict variability for unclassified stars

Data Overview

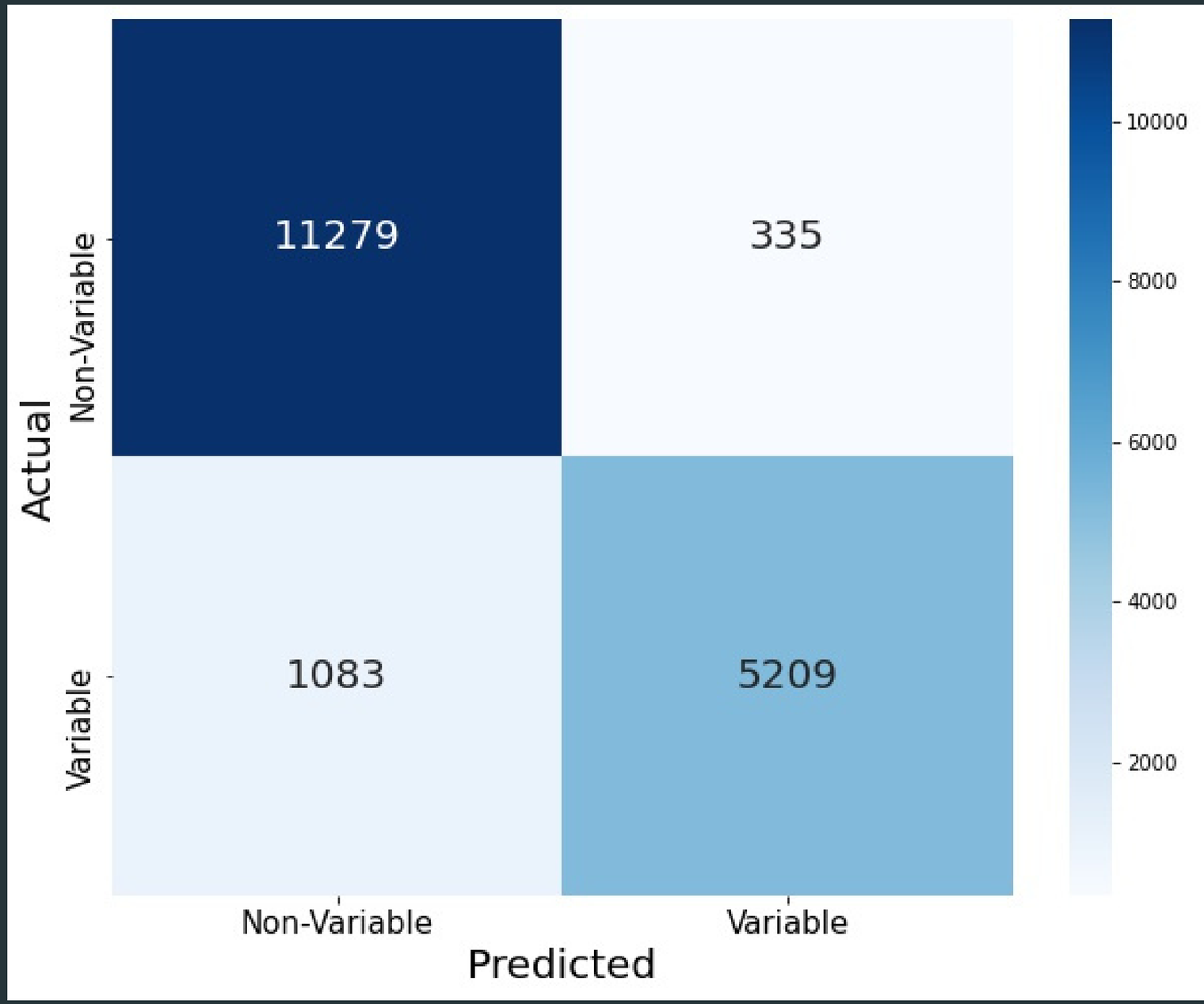
- Used data gathered from ESA's Hipparcos Satellite
- Hipparcos was active from 1989 to 1993
- Data has 78 features and lists 118,218 stars

Data Handling

- Target had values indicating unclassified stars
- Separated into two datasets
- Dataset to model on, dataset to predict on
- Data needed to be cleaned in multiple different ways

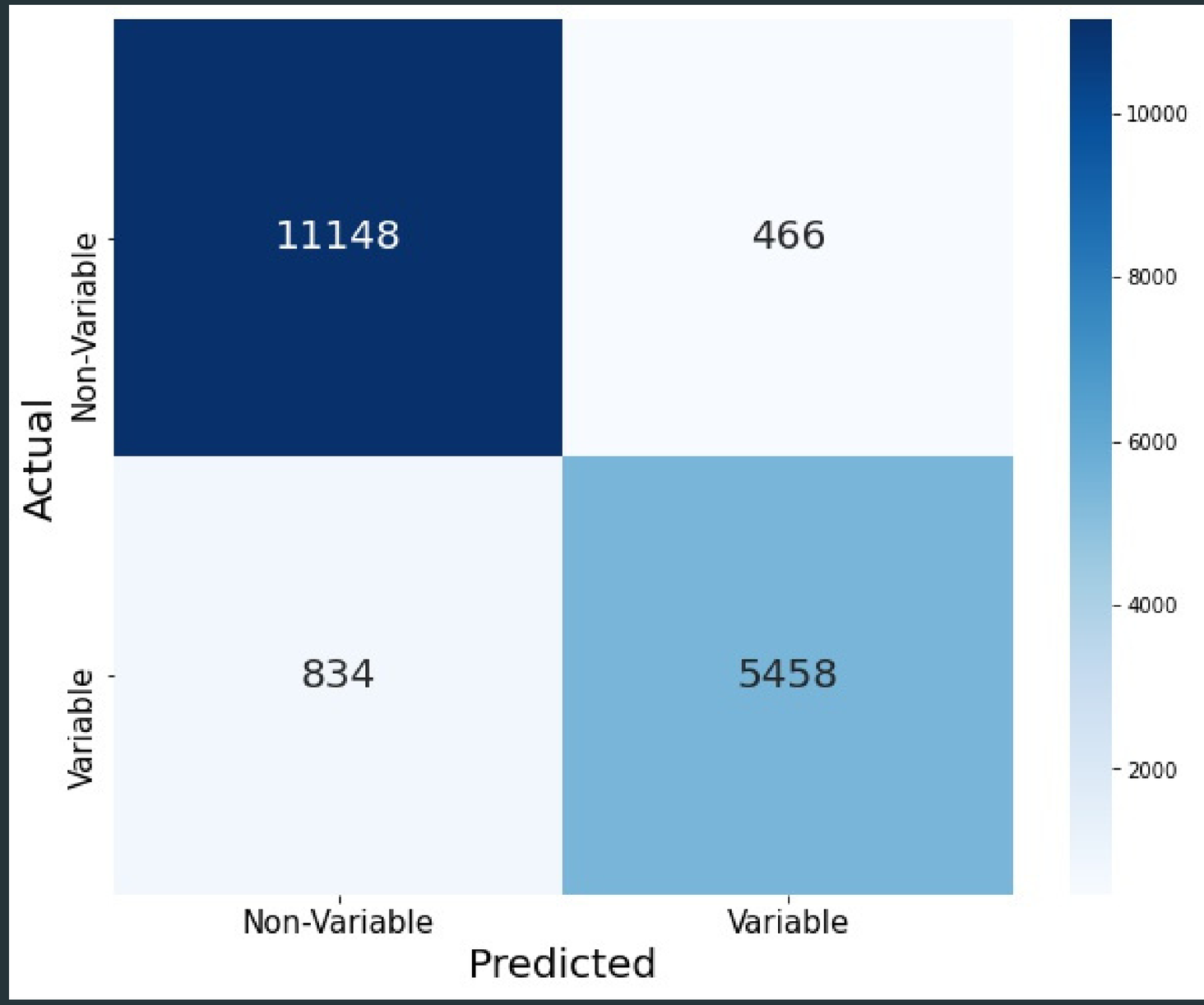
Initial Model

- Model using all features
- Many false negatives
- Few false positives
- Overall quite accurate



Final Model

- Smarter about which data was used
- 9 columns used
- More accurate overall than our other models
- Fewer false negatives



Results

- Model predicted variability of unclassified stars
- Predictions show which stars are expected to be variable
- We have a list of stars predicted as variable

Next Steps

Further Explore Data

More Model Tuning

Further Reduce False Negatives


Any
Questions?