Informing a Move

The Data Science Way

Mark Ehler

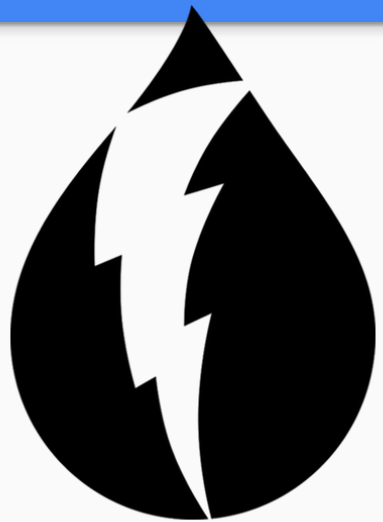
Table of Contents

The Problem at Hand

Commercialization of NOAA data

5G and polar passive microwave satellites


Temperature in F
80
60
40
20
0
−20
980
990
1000
1010
1020
1030
1040
Pressure in Pa
Clear
Cloudy
Mostly Cloudy
Overcast
Partly Cloudy
Rain
Snow



Powered by Dark Sky

The **Dark Sky** API is backed by a wide range of global weather **data sources**, which are aggregated together to provide the most accurate forecast possible for a given location. Any **data sources** used to service a given request will be noted in the flags section of the API response.

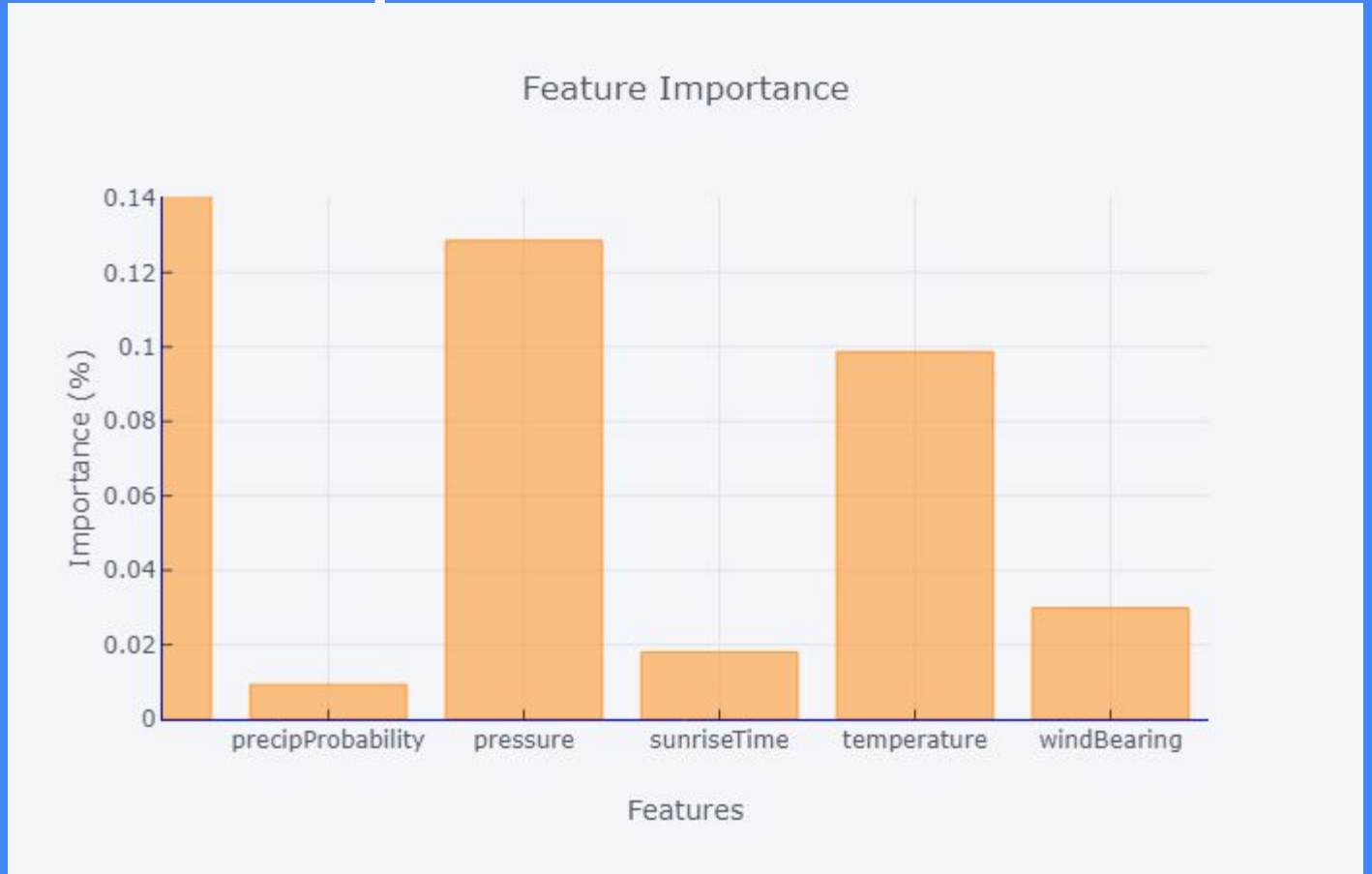
Feature Engineering

Change in weather - inferred from clever API calls

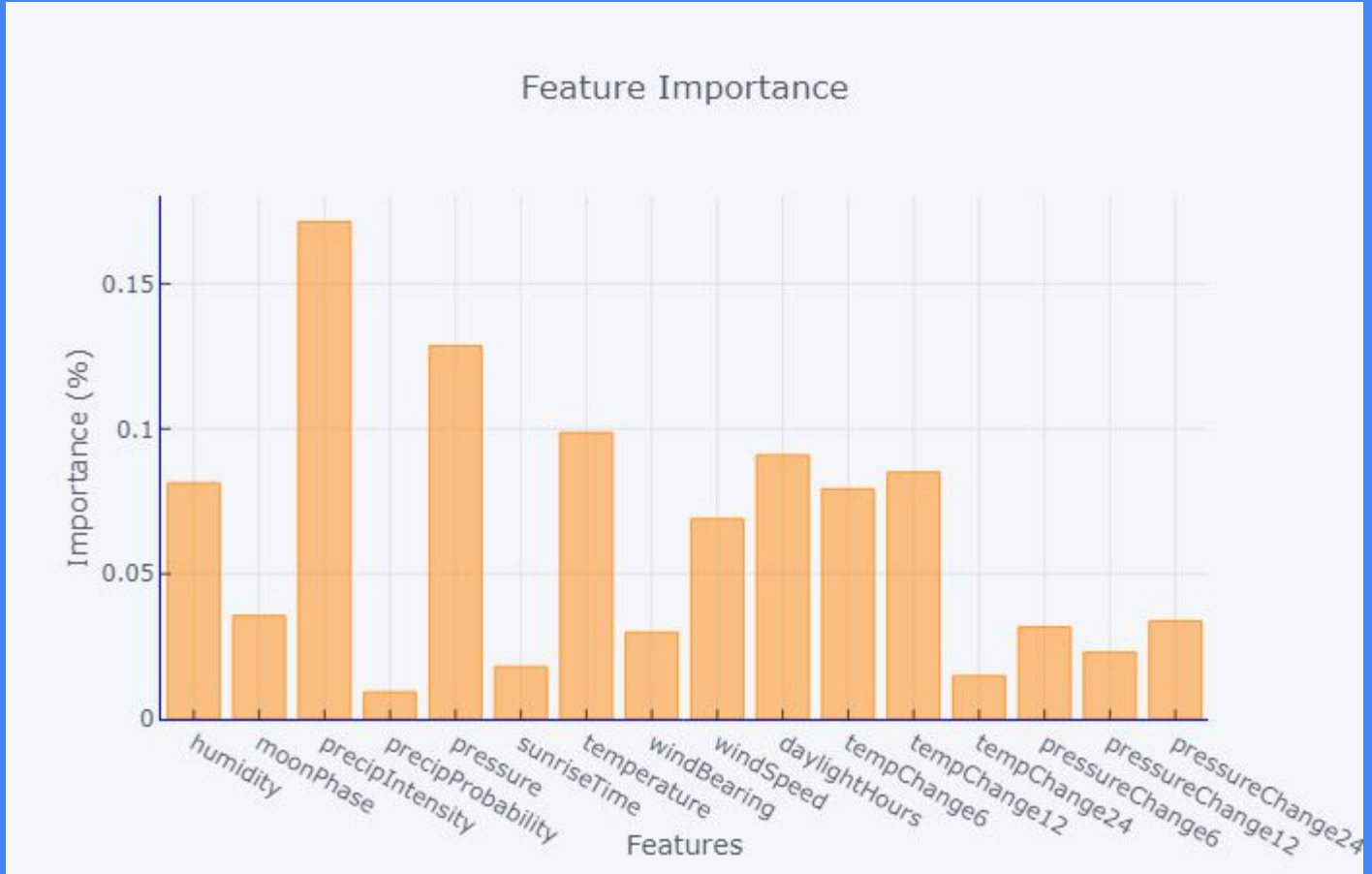
Seasonal changes - daylight hours, moon phase

Removing wrong minded features - facts don’t always do what you want them to. Guardian of 21st century definition of what is truth.

Feature Importance



Feature Importance

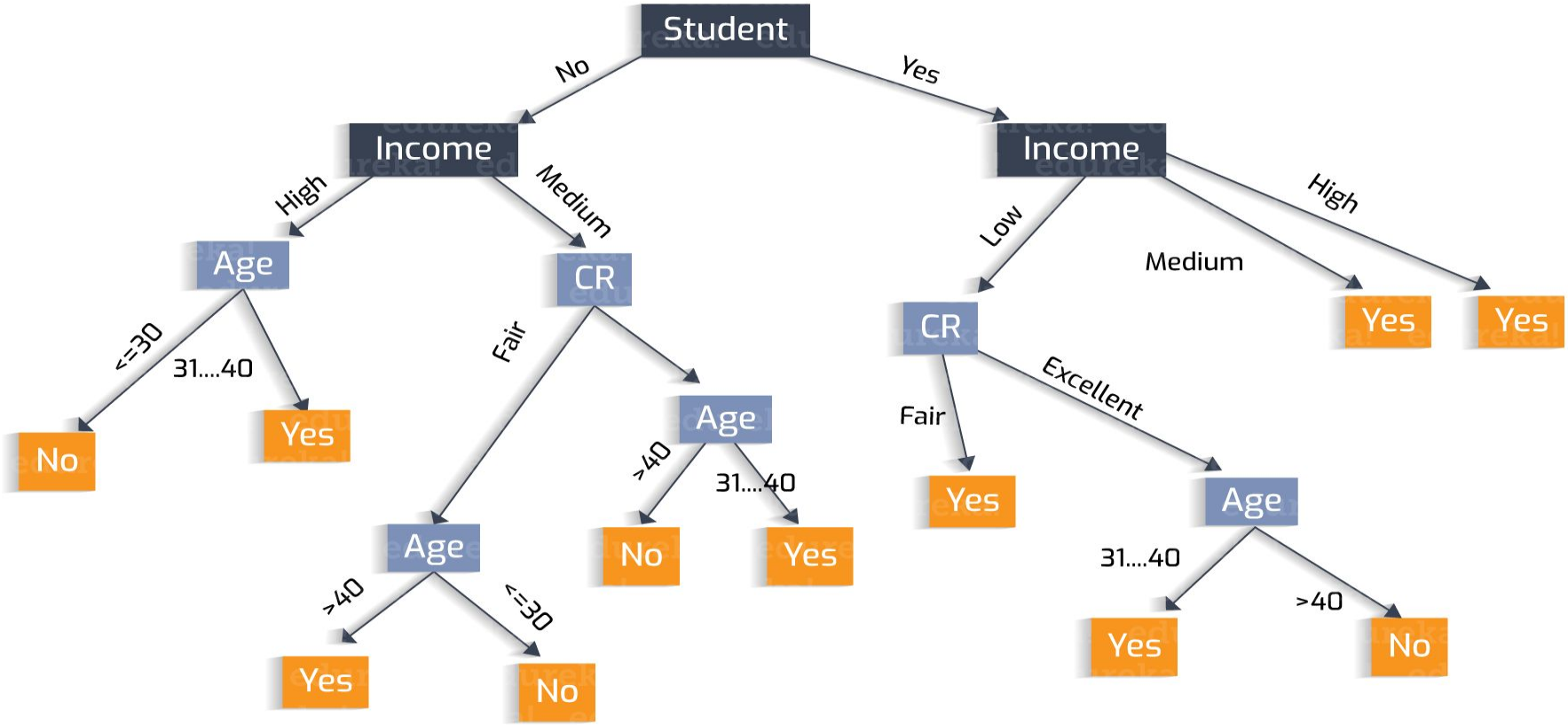


Decision Tree

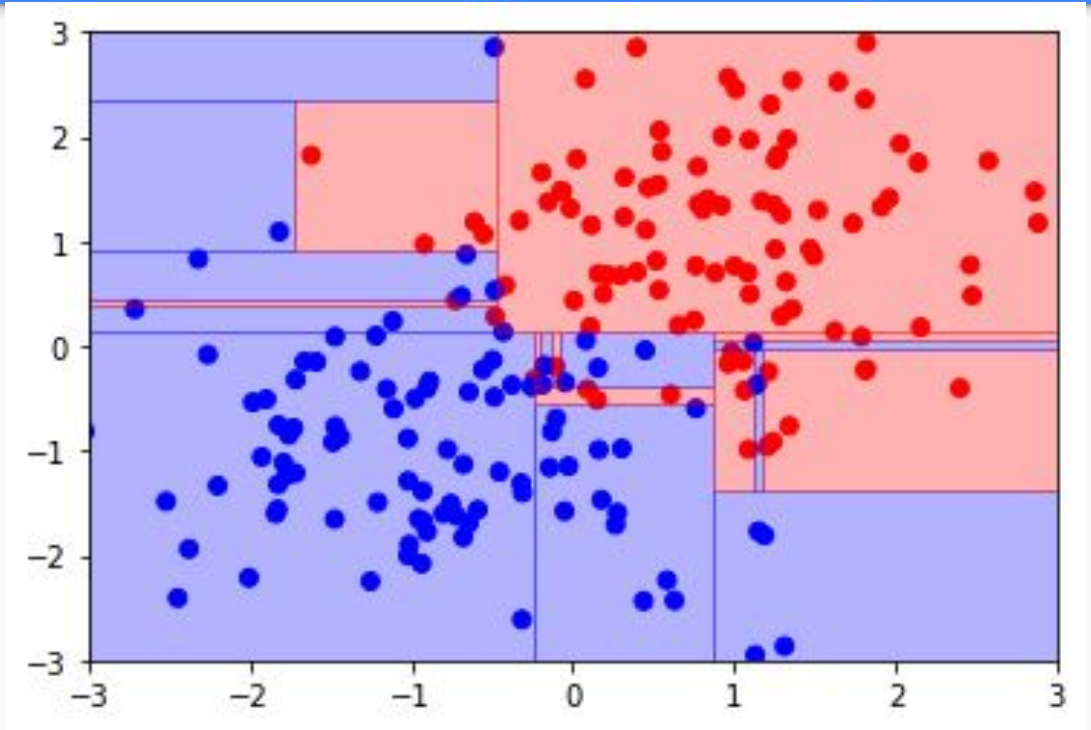
A classic and still-relevant way for computers to categorize and make decisions. Logical and linear learning with accessible readability.

At any time, we can check and see what the computer is thinking.

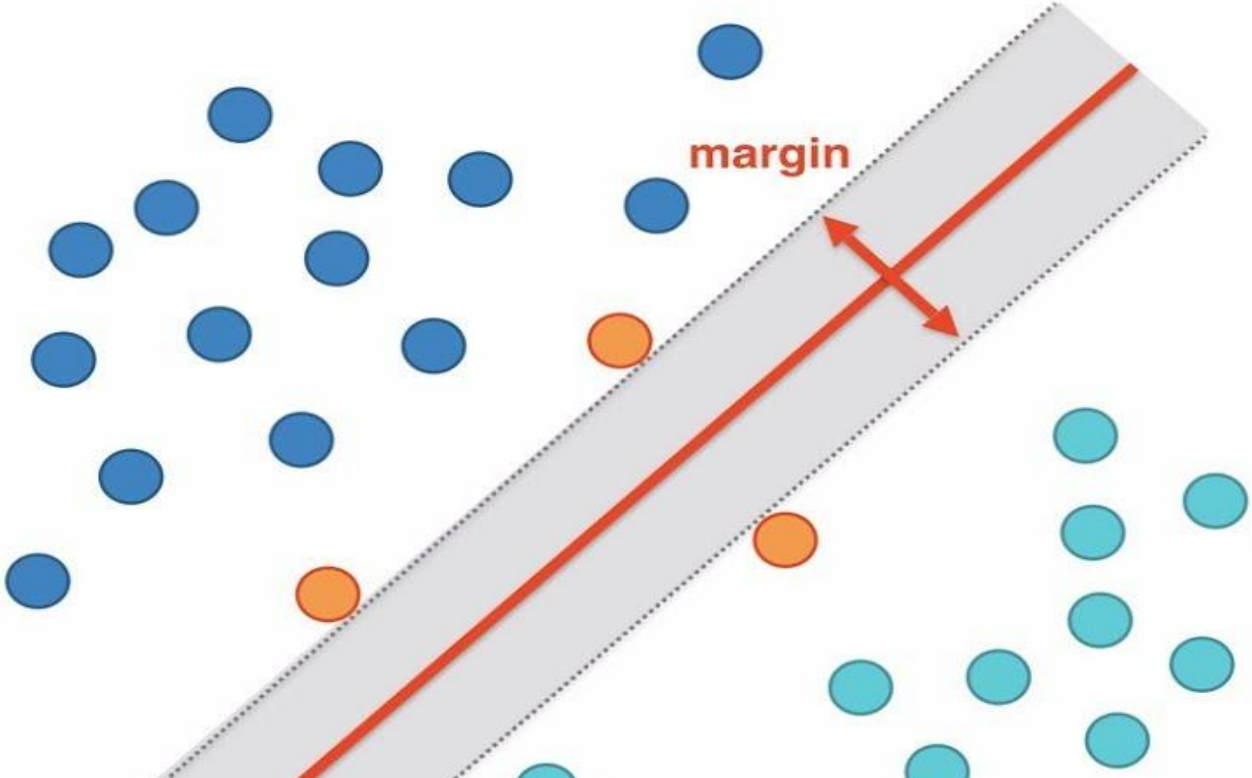
Decision Tree - Credit Card Approval (example)



SVM vs Decision Tree



SVM - Hyperparameters



Why is this the best split?

The distance between the support vectors and the hyperplane are as far as possible

Random Forest

Results 4.0

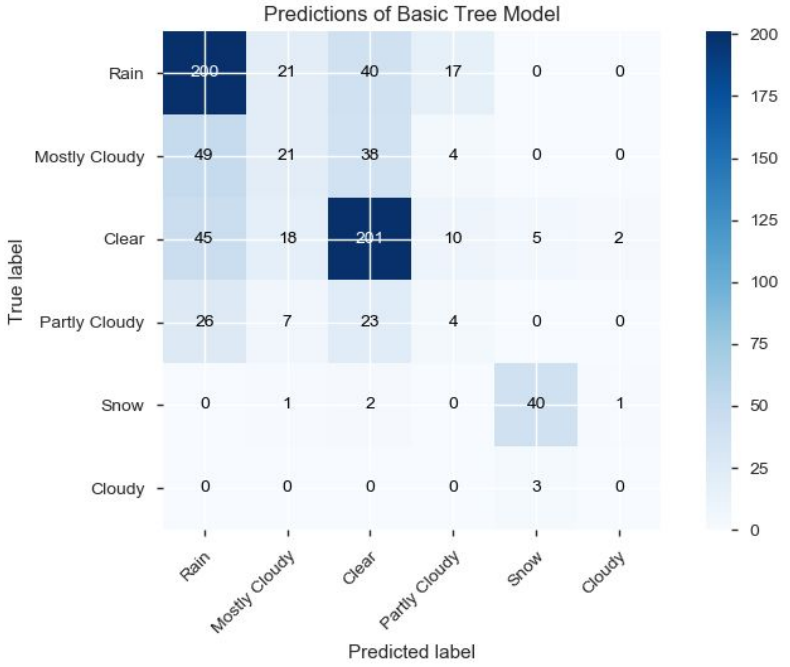
Training Accuracy: 76.05%

Validation Accuracy: 67.47%

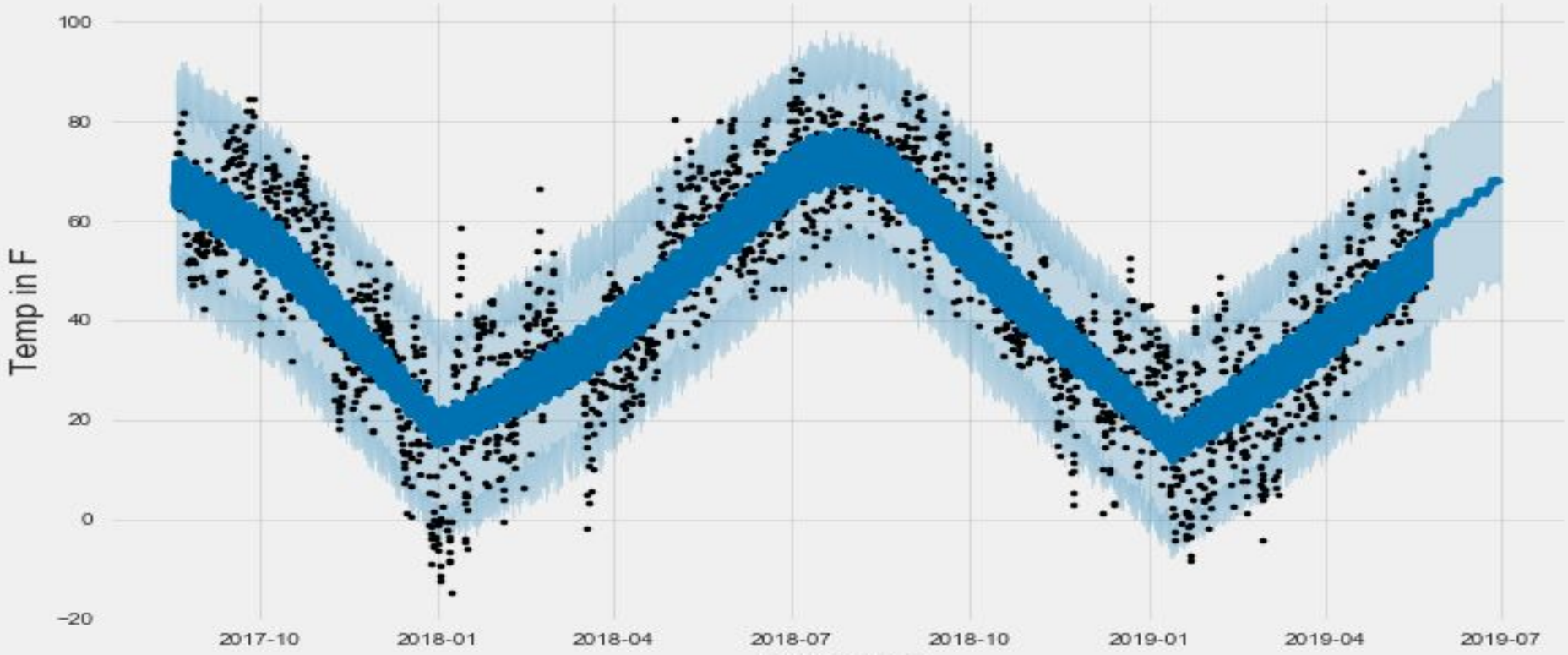
A Word on Accuracy

- Accuracy defined as # of misses vs correctly categorized weather
- Recall as # of misclassifications

How do we want our errors to come? As missed storms or false alarms?

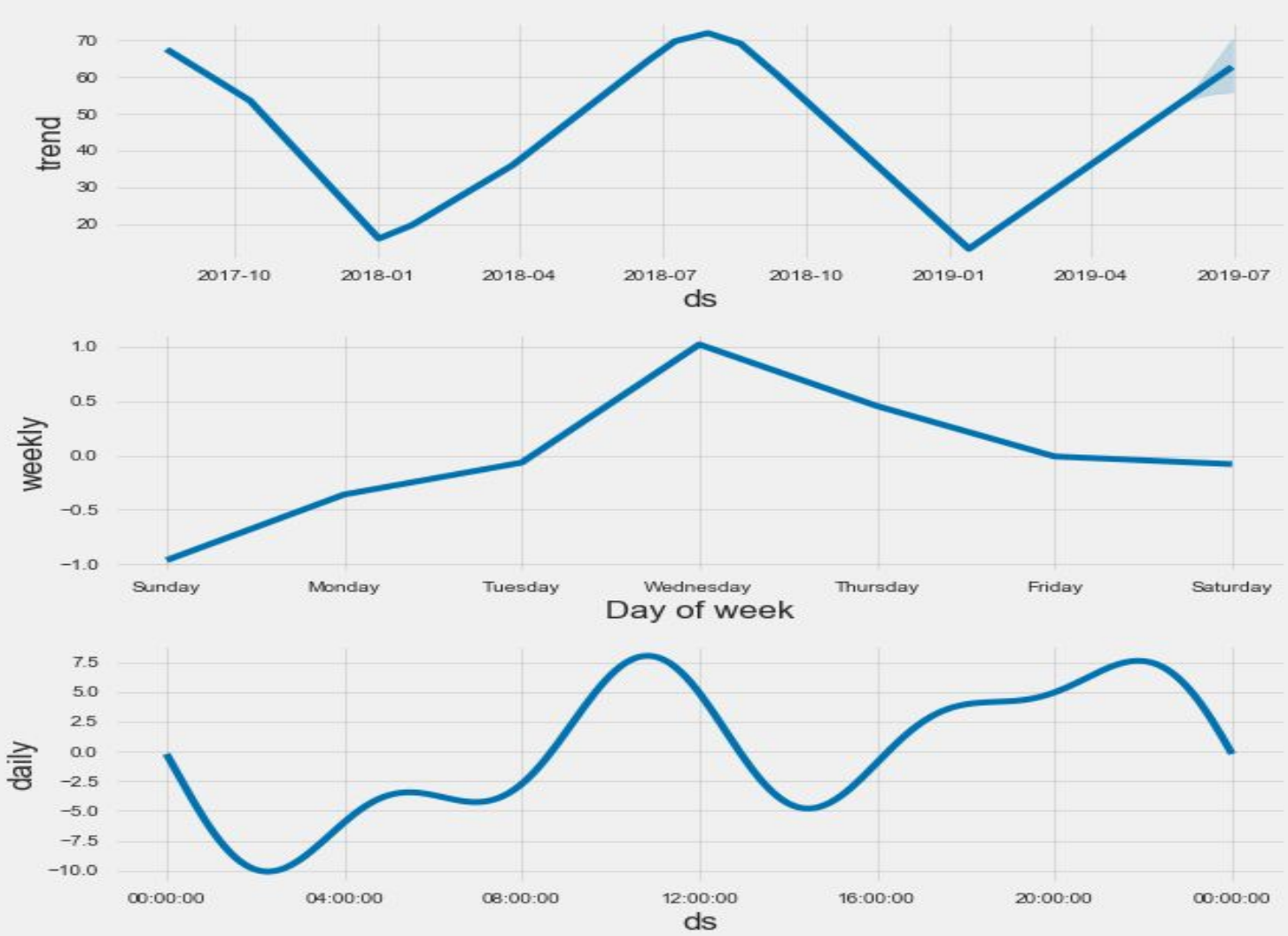


Predicting Temperature



Finding Trends

How can we use this to improve our model?



Further Research

Upsampling precipitation events.

Collecting data from surrounding stations.

Using the entire dataset to predict future outlooks.

Appendix

Linkedin

Github Repo

Blog

Washington Post Article

A Visual Introduction to Machine Learning