Group ID: 3

PREDICTING HOTSPOT INTENSITIES

Say Yueyang, Symus | He Zeqing | Kwek Yan Qing

TABLE OF CONTENTS

INTRODUCTION

Indonesia's Forest Fires

Haze

MOTIVATION

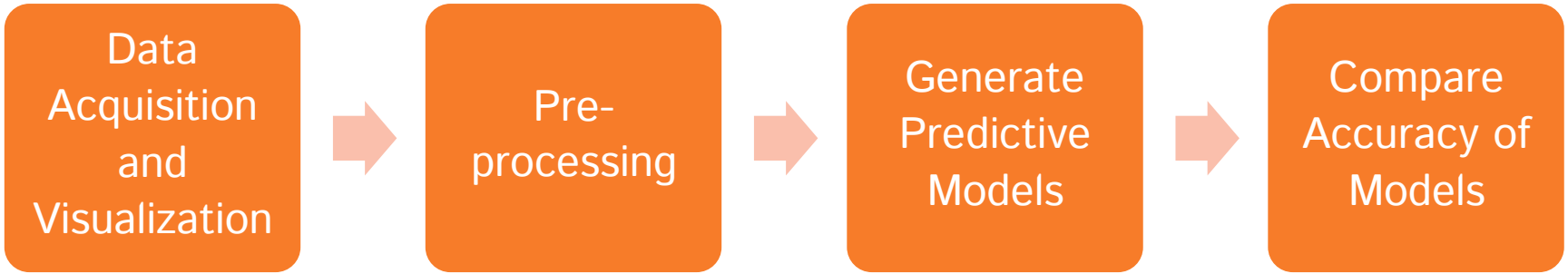
Brightness data

Deploy firefighting teams

Prevent fires from escalating

Can we predict **how intense a fire will be** based on **where it starts** and **how bright the fire burns**?

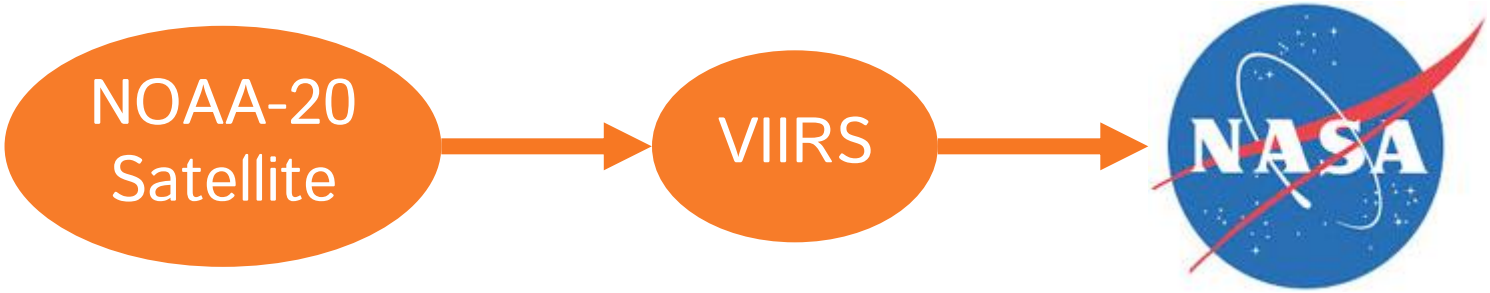
PROPOSED SOLUTION



To predict **Fire Radiative Power (FRP)** from location and brightness data.

PROPOSED SOLUTION

Data Acquisition

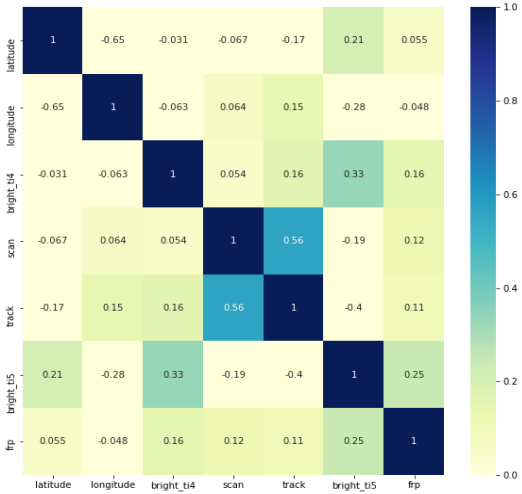


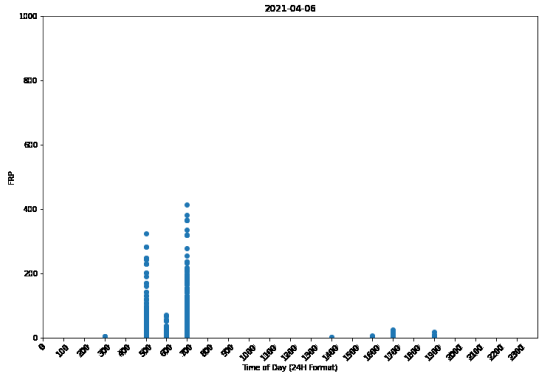
Data extracted from NASA's Fire Information for Resource Management System (FIRMS).

- Dynamic data; near real-time (3h delay)
- Dataset available: 24h/48h/**7 days**

PROPOSED SOLUTION

Data Visualisation





Seaborn

Correlation Matrix, Scatter Plot

Plotly

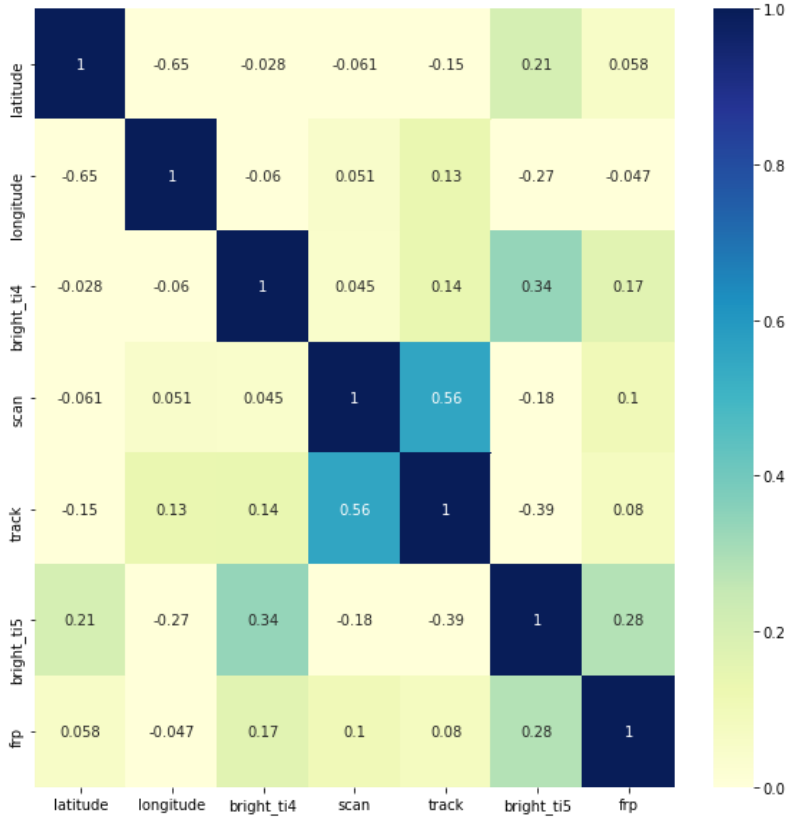
Time-Series

Matplotlib

Time-Series (Hourly)

PROPOSED SOLUTION

Data Visualisation



Correlation Matrix

- Observed correlation between Brightness and FRP
- Observed correlation between scan and track (related to satellite movement)

PROPOSED SOLUTION

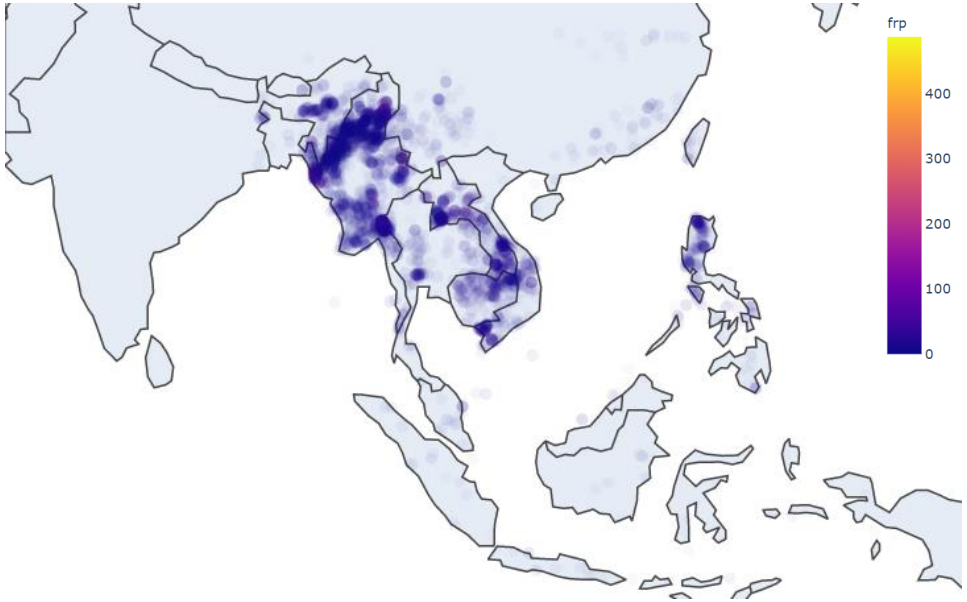
Data Visualisation

Time-Series Plot

- Plot of FRP values for **every satellite recording**
- Observed repeated/segmented data

PROPOSED SOLUTION

Data Visualisation

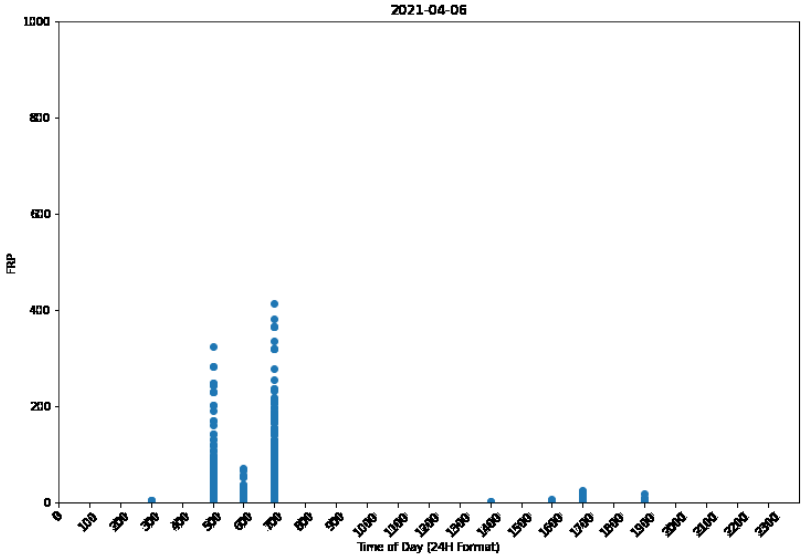


Time-Series Plot

- Plot of FRP values animated **over every day**
- Observed concentration of fire data in same region

PROPOSED SOLUTION

Data Visualisation



Time-Series Plot

- Plot of FRP values for **every hour, animated over every day**
- Observed peaks at certain times of day

DATA PRE-PROCESSING

Reformatting Features & Processing The Dataset

One Hot Encoding (OHE)

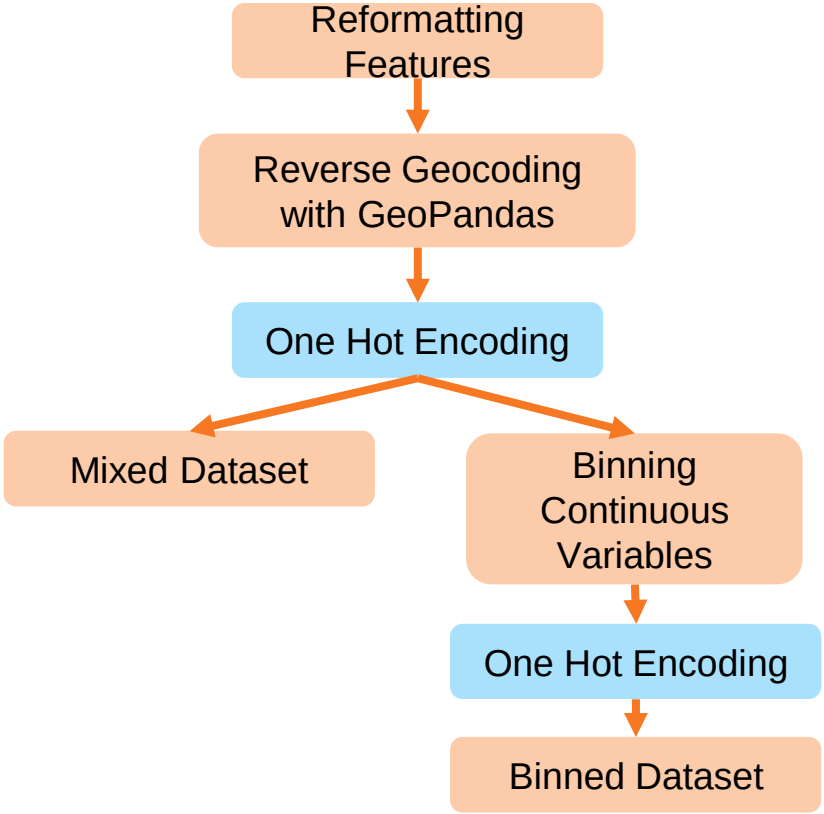
Confidence

Confidence_low
Confidence_nominal
Confidence_high

Daynight

Daynight_D
Daynight_N

Values produced: 0,1



DATA PRE-PROCESSING

Reformatting Features & Processing The Dataset

Dataset #1 – Conversion of Location to Binary Values

continent	name
Oceania	Papua New Guinea
Oceania	Papua New Guinea
Oceania	Papua New Guinea
Asia	Philippines
Asia	Philippines

OHE

...	name_Indonesia	name_Laos	name_Malaysia	name_Myanmar	name_Nepal	name_Papua New Guinea	name_Philippines	name_Taiwan	name_Thailand	name_Vietnam
...	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
...	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
...	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
...	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
...	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0

Dataset #2 – Conversion of Continuous Values to Binary Values

bright_ti4	scan	track	acq_time	bright_ti5
330.7	0.43	0.38	33000.0	293.1
328.9	0.53	0.50	33000.0	289.9
326.2	0.52	0.50	33000.0	287.4
335.1	0.41	0.37	51200.0	298.4
336.0	0.41	0.37	51200.0	298.0

OHE

bright_ti4_bright_ti4_100p	bright_ti4_bright_ti4_20p	bright_ti4_bright_ti4_40p	bright_ti4_bright_ti4_60p	bright_ti4_bright_ti4_80p	scan_scan_100p	scan_scan_
0.0	1.0	0.0	0.0	0.0	0.0	
0.0	1.0	0.0	0.0	0.0	1.0	
0.0	1.0	0.0	0.0	0.0	0.0	
0.0	0.0	1.0	0.0	0.0	0.0	
0.0	0.0	0.0	1.0	0.0	0.0	

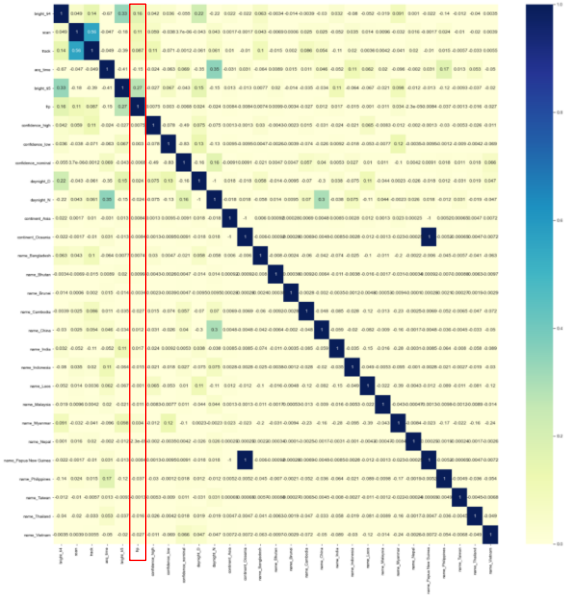
DATA PRE-PROCESSING

Feature Selection

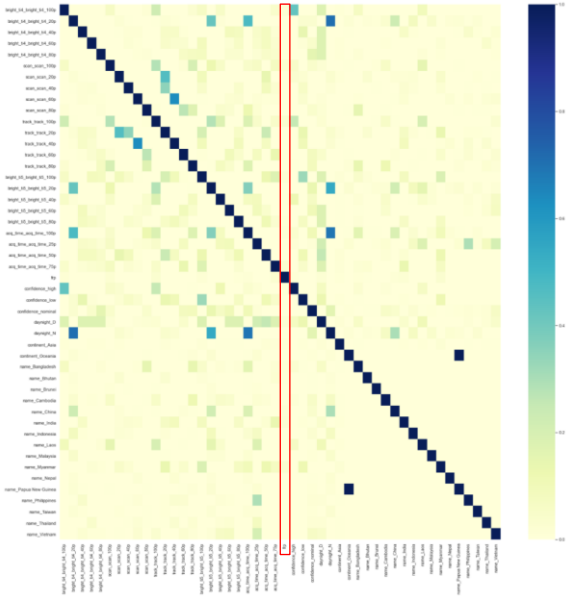
Correlation Matrix + Heatmap

```
from sklearn.preprocessing import StandardScaler
def standardize(df):
  # create a scaler object
  std_scaler = StandardScaler()
  # fit and transform the data
  return pd.DataFrame(std_scaler.fit_transform(df), columns=df.columns)
```

Continuous + Binary Values



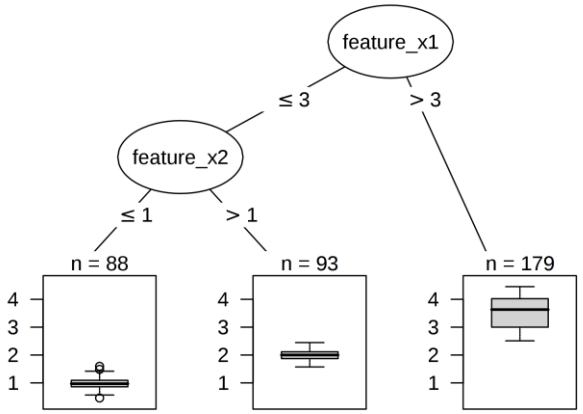
Binary Values

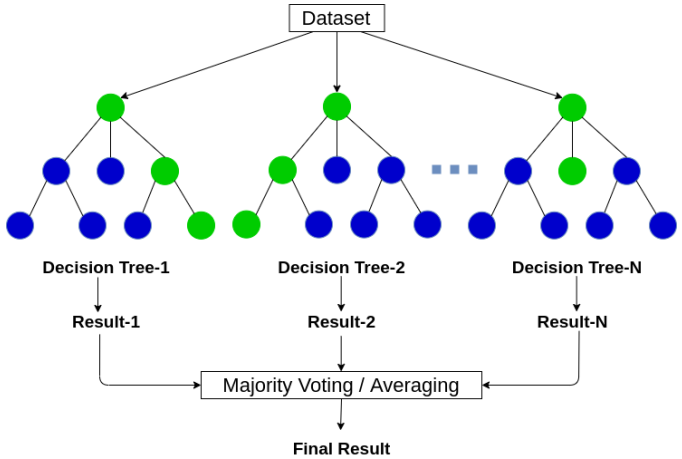


DATA PRE-PROCESSING

Feature Selection

Random Forest Feature Selection Model





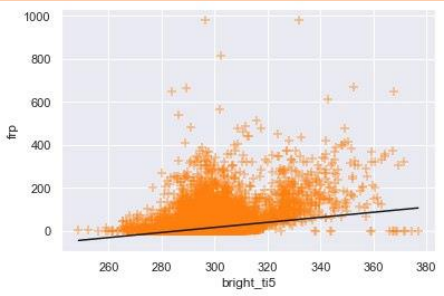
The top 3 factors correlating to the Fire Radiative Power are bright_ti5_bright_ti5_40p, bright_ti5_bright_ti5_80p, bright_ti4_bright_ti4_20p

The top 3 factors correlating to the Fire Radiative Power are bright_ti5, bright_ti4, scan

PREDICTIVE MODELLING

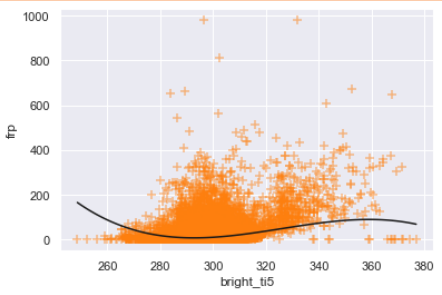
Types of Models

Linear Regression



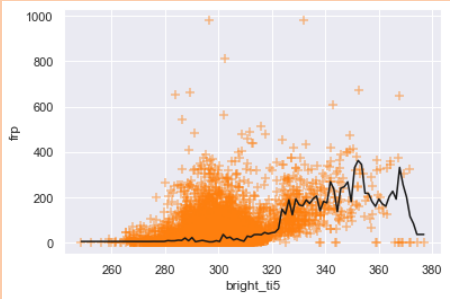
$$FRP = \Sigma w_n x_n + w_0$$
$$= \mathbf{Xw}$$

Poly Regression



$$FRP = \Sigma(\Sigma w_{n,m..} x_n x_m ..)_d$$
$$+\Sigma w_n x_n + w_0$$

Random Forest



$FRP =$
mean of n (20) decision trees outputs

PREDICTIVE MODELLING

Train-Val-Test Split

Linear Model

Random Forest Model

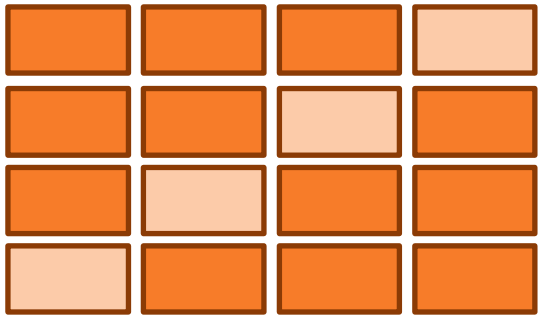
Train Set

Test Set

Model parameters trained using **Train Set**

Evaluate Accuracy using **Test Set**

Polynomial Model

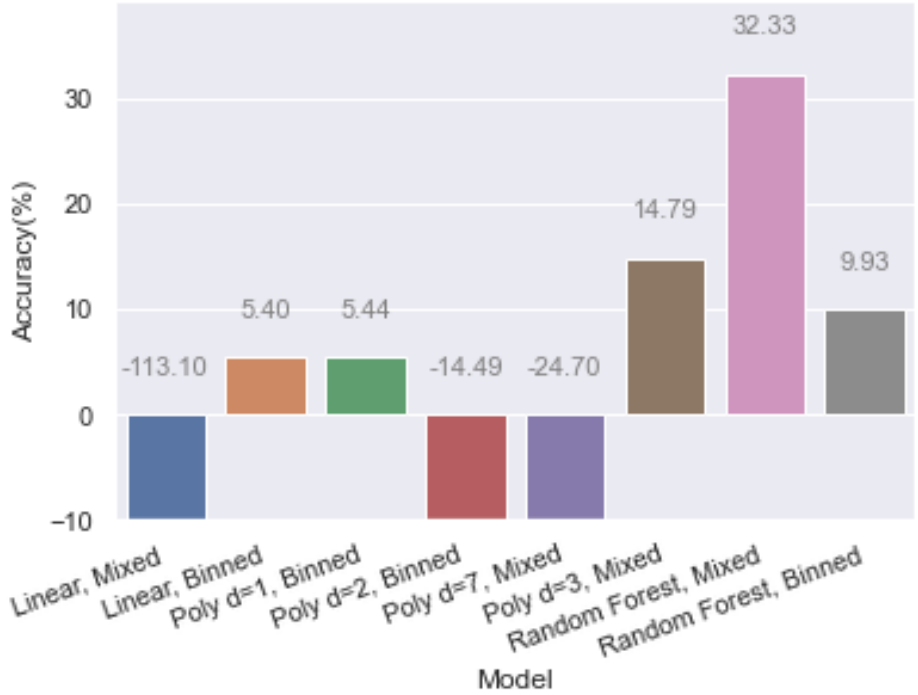


Test Set

Train + Validation Set

4-fold Cross Validation:
Find ideal degree *d*
(hyperparameter)

RESULTS AND ANALYSIS



$$\text{Accuracy} = 1 - \frac{1}{n}\sum\frac{|Error|}{|Actual|}$$

Random Forest Model captures the non-linear trends in the data (Best Accuracy: **33%)**

Conclusion: There is not much trend between location or brightness and FRP

FUTURE IMPROVEMENTS

Find dataset with **better** features

Use more data over **longer** periods

Reduce the problem into classifying high FRP values versus low FRP values

Identify locations of interest and analyse location characteristics **over a smaller area**

THANK YOU!

This project is available on GitHub!
Click on the logo to view it!

CREDITS: This presentation template was modified from a template by Slidesgo, including icons by Flaticon, and infographics & images by Freepik

CONTRIBUTIONS

Symus
Data Exploration
Data Visualization
Data Cleaning
One Hot Encoding

Atticus
Data Cleaning
One Hot Encoding
Feature Selection
Contextual Knowledge and Analysis
Formatting

Zeqing
Creating Models
Cross Validation
Evaluating Model Accuracy