DS3000 Final Project

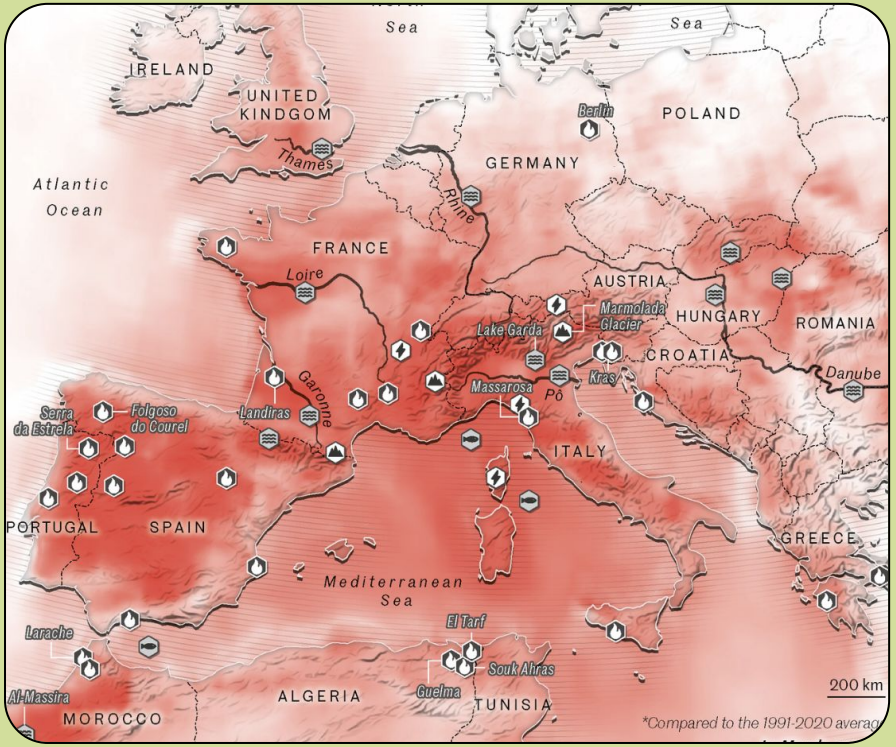
Predicting Wildfire Trends in Europe Using Machine Learning

Michael Hrinda | Elana von der Heyden | Anya Wild | Janina Kurowski

Problem Statement/ Goal

Problem Statement: Wildfires in Europe are increasing in frequency and intensity. Threats include biodiversity loss, economic damage, and human displacement.

Goal: Use ML to predict wildfire trends to guide preventive actions.



Significance

- EU Forest Strategy for 2030 focuses on prevention but lacks predictive capabilities.
- NASA’s satellite data and ML offer predictive opportunities.

Data Description

Source

NASA's Fire Information for Resource Management System.

Dataset

Visible Infrared Imaging Radiometer Suite (VIIRS) 375m for high quality image resolution.

Key Attributes

- Brightness: intensity of fire
  - I-4 Channel (high)
  - I-5 Channel (low)
- Longitude and Latitude
- Date and Time of Image Acquisition
- Scan and Track: image resolution in each direction

Data Cleaning

Steps:

- FIRMS API for specifically European countries and relevant dates
- Filtered for target attributes
- Wrote CSV file for easier access

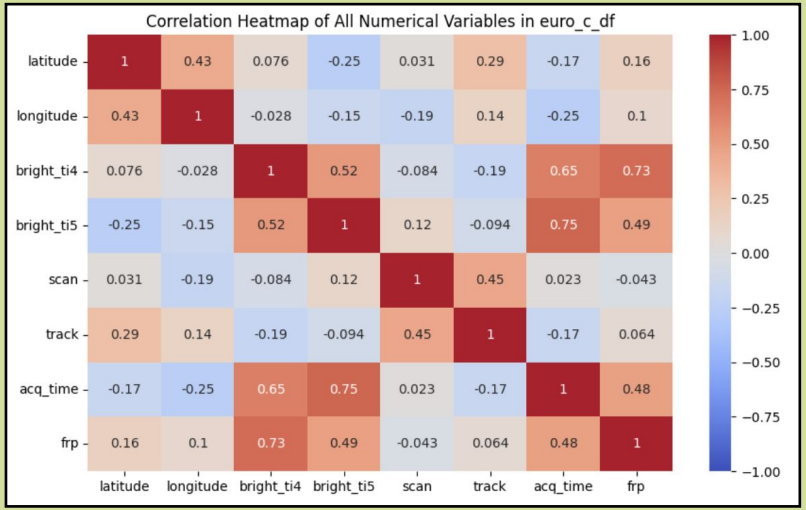
Clean Data Frame:

country_id	latitude	longitude	bright_ti4	bright_ti5
AUT	47.34311	9.62378	328.5	290.2
AUT	47.54527	9.78854	331.0	287.0
AUT	47.54559	9.78841	329.3	286.5
AUT	48.27758	14.34202	331.8	276.3
AUT	48.27502	14.33618	300.4	279.2

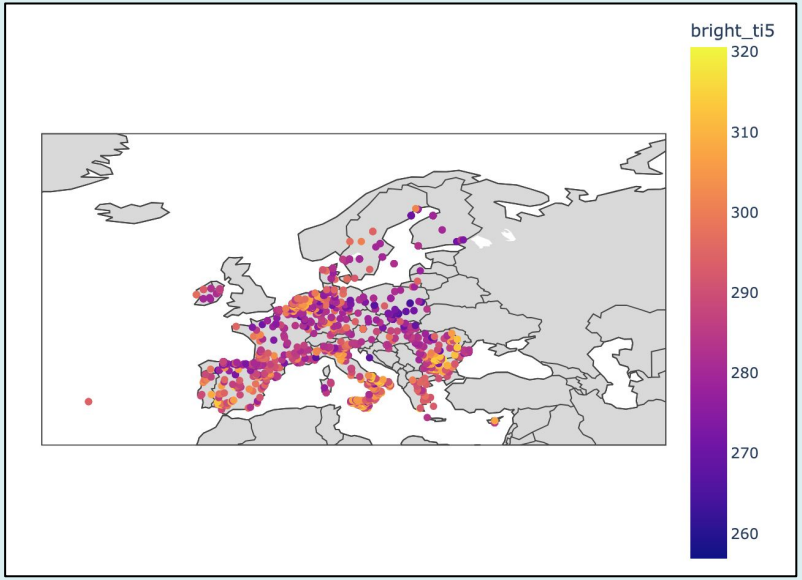
scan	track	confidence	acq_date	acq_time	frp	daynight
0.53	0.42	n	2020-03-01	1230	4.4	D
0.54	0.42	n	2020-03-01	1230	5.0	D
0.54	0.42	n	2020-03-01	1230	3.7	D
0.59	0.53	n	2020-03-01	1230	3.8	D
0.58	0.70	n	2020-06-01	218	1.9	N

Visualizations

Correlation Heatmap of Numerical Attr.



Heatmap of I-5 Channel Brightness



Linear Regression

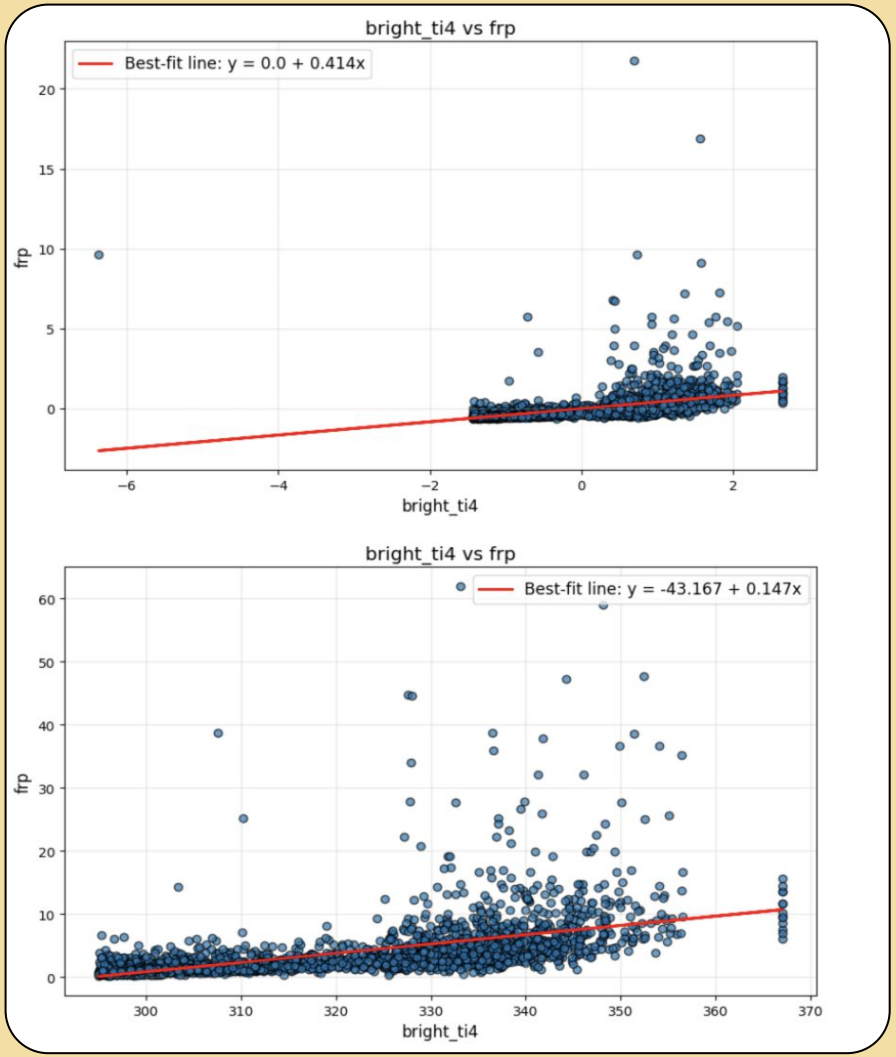
Model 1

Objective: Predict FRP based on bright_ti4.

Results:

- R^2 = 0.171, MSE = 0.829.
- Issues with outliers and assumption violations.

Visualization: Scatter plot of FRP vs. bright_ti4 (before and after removing outliers).



Model 2

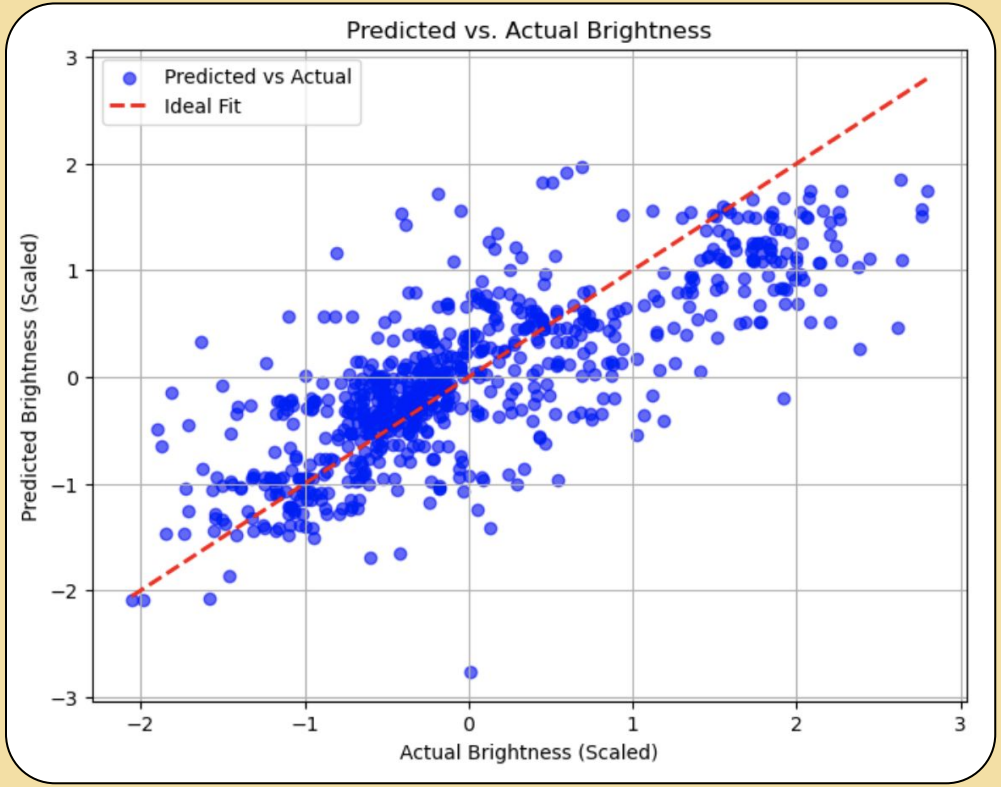
Polynomial Regression

Objective: Predict bright_ti5 using date, latitude, longitude, scan, and track.

Results:

- R^2 = 0.374, MSE = 0.626 (improved from linear regression).
- Cross-validation: Consistent results, minimal overfitting.

Visualization: Predicted vs. actual bright_ti5 values.



Classification: Logistic Regression

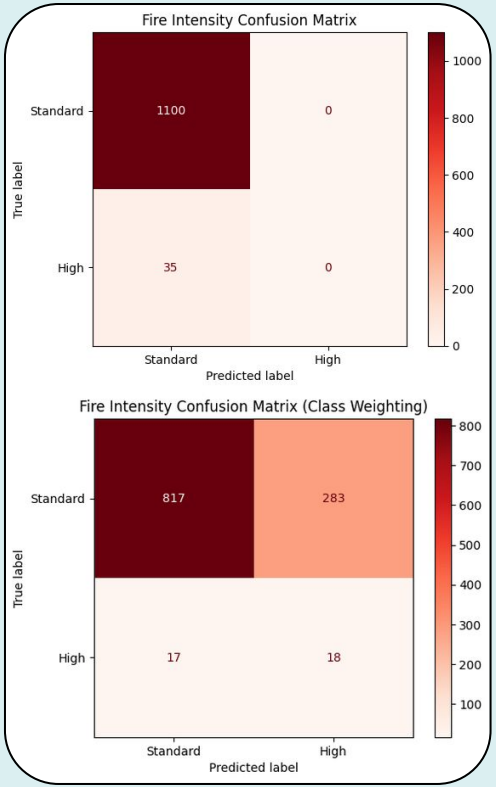
Model 3

Objective: Predict fire intensity (standard or high) given date, latitude, longitude, scan, and track as features

Results:

- Logistic Regression without Balanced Class Weighting:
  - Accuracy: 0.969 Precision: 0.000 Recall: 0.000 F1-Score: 0.000
  - AUC-ROC: 0.691
- Logistic Regression with Balanced Class Weighting:
  - Accuracy: 0.736 Precision: 0.060 Recall: 0.514 F1-Score: 0.107 AUC-ROC: 0.689

Visualization: Confusion matrices



Limitations and Future Directions

Limitations

- Our ‘best’ model only explained 37% of the variance
- Limitations of our dataset and the FIRMS API

Future Directions

- Include additional variables such as weather patterns
- Test with a new region such as the U.S.
- Explore advanced algorithms (e.g., neural networks)