Machine Learning Analysis of Factors Influencing Wildfire Distribution

Geoinformatics Project

By: Feras Alqrinawi

Jul. 8 , 2024

Supervised by: Prof. Venuti Giovanna ,
Dr. Daniela Stroppiana

POLITECNICO
MILANO 1863

Introduction – Study Area

Study Area: The Seven Continents

•**Africa:** Frequent wildfires in **savannas** and forests due to climate, vegetation, and human activities.

•**Asia:** Diverse ecosystems with wildfires influenced by land use, deforestation, and agriculture.

•**Europe:** Mainly in the **Mediterranean**, driven by hot, dry summers and human activities.

•**North America:** Severe wildfires in the western regions due to climate change, forest management, and population growth.

•**Australia:** Prone to **bushfires** due to hot, dry conditions and flammable vegetation.

•**South America:** Significant fires in the **Amazon**, driven by deforestation and agricultural expansion.

•**Antarctica:** Rare wildfires, but climate change is causing potential future risks.

Objectives

Study Area: The Seven Continents

•**Africa:** Frequent wildfires in savannas and forests due to climate, vegetation, and human activities.

•**Antarctica:** Rare wildfires, but climate change is causing

Objective 1: Determine environmental and human-related factors that significantly influence wildfire distribution.

Objective 2: Create models to predict wildfire risks based on identified factors

Objective 3: Identify the factors with the greatest effect on wildfire occurrences.

•Uses comprehensive data to understand multifaceted factors influencing wildfires

Methodology - Data Collection

Study Area: The Seven Continents

•**Africa:** Frequent wildfires in savannas and forests due to climate, vegetation, and human activities.

•**Antarctica:** Rare wildfires, but climate change is causing

After searching and gathering data from a variety of <u>reliable and relevant sources</u>[1].

We got :

Environmental Factors

- Climate data (precipitation, relative humidity, solar radiation).
- Vegetation indices (NDVI, NDMI).
- Topographic data (slope, aspect, elevation).
- Soil moisture and evapotranspiration rates.

Environmental Factor	Description
Precipitation	The amount of rainfall or snowfall, measured in **millimeters**.
Relative Humidity	The amount of moisture in the air compared to what the air can hold at that temperature, expressed as a **percentage**.
Solar Radiation	The amount of solar energy received by a specific area, measured in **watts per square meter**.
NDMI	Normalized Difference Moisture Index, used to determine vegetation water content. **((Band 5 – Band 6) / (Band 5 + Band 6))**
NDVI	Normalized Difference Vegetation Index, used to assess whether the target being observed contains live green vegetation. **(Band 5 – Band 4) / (Band 5 + Band 4))**
Slope	The steepness or incline of the land, usually expressed as a **percentage**.
Aspect	The compass direction that a slope faces, usually measured in degrees from north. **(Categorical)**
Elevation	The height of the land above sea level, measured in **meters**.
Soil Moisture	Soil moisture content, indicating the amount of water contained in the soil.
evapotranspiration	The sum of evaporation and plant transpiration from the Earth's land and ocean surface to the atmosphere.
Land Cover	The physical material at the surface of the earth, such as vegetation, urban infrastructure, water, etc**. (categorical)**

After searching and gathering data from a variety of reliable and relevant sources[2].

We got :

Human-Related Factors and :

- Human Impact Index.
- Population density.

other Factors:

- Fuel Load
- Fuel Moisture Content
- Lightning Frequency

Imported and processed each image collection that has these

Factor	Description
Human Impact Index	A measure of human impact on the environment, averaged over a specific area. **(Numerical)**
Population Density	The number of people living per unit of area, usually measured in people per square kilometer. **(people/km2)**
Fuel Load	Global Aboveground and Belowground Biomass Carbon Density **measured in tons per hectare.**
Fuel Moisture Content	The amount of moisture in the fuel, expressed as a **percentage**.
Lightning Frequency	The frequency of lightning strikes in a given area, measured in **strikes per year**.

- .median() function is applied to the processed image collection of Landsat

Slope , Aspect	NASA SRTM Digital Elevation 30m	• No Preprocessing (Data and are static over short periods, no aggregation was applied.)

Methodology - Data Preprocessing

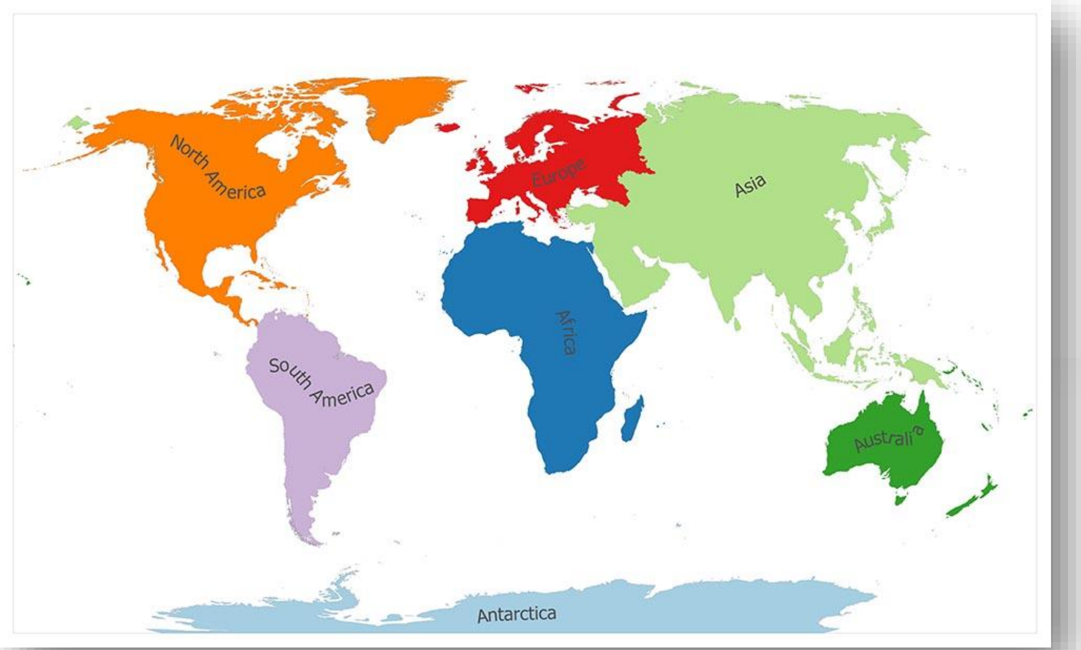
The Study Area is really big, so we need to take some steps to make it easier for Google Earth Engine to process and get the results we need, so:

*** If we don't follow these steps, we will run into runtime errors because Google Earth Engine has limited resources.**

Methodology - Data Preprocessing

The Study Area is really big, so we need to take some steps to make it easier for Google Earth Engine to process and get the results we need, so:

1- Clip the study area **by continent**, assigning each continent its own shapefile to process individually and exclude the oceans. **This approach reduces computational strain and speeds up processing.**



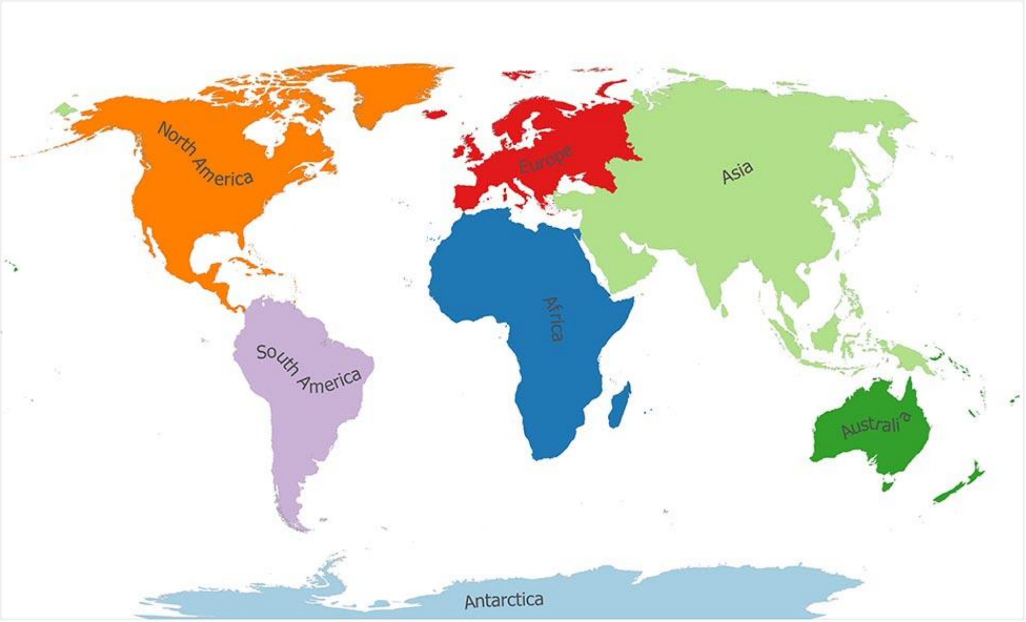
* If we don't follow these steps, we will run into runtime errors because Google Earth Engine has limited resources.

Methodology - Data Preprocessing

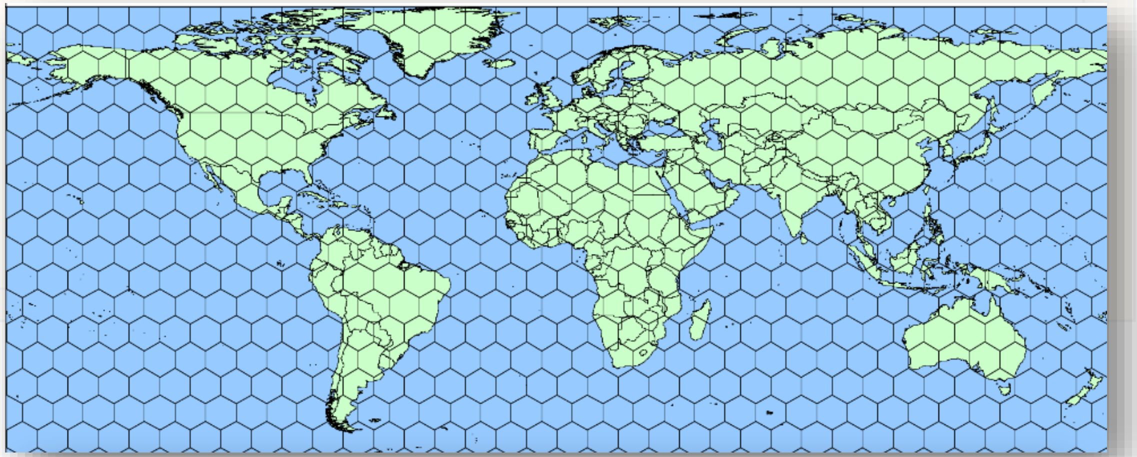
The Study Area is really big, so we need to take some steps to make it easier for Google Earth Engine to process and get the results we need, so:

* If we don't follow these steps, we will run into runtime errors because Google Earth Engine has limited resources.

1- Clip the study area **by continent**, assigning each continent its own shapefile to process individually and exclude the oceans. **This approach reduces computational strain and speeds up processing.**



2- Use QGIS to create **500 km** (horizontal/vertical spacing) **hexagon grids**, representing geographic areas on a broader scale, and upload them to Google Earth Engine as an asset. This helps **visualize data with a lower level of detail.**



Methodology - Data Preprocessing

Imported and processed each image collection that has these factors Using Google Earth Engine here some :

Environmental Factor	Image Collection	Preprocessed ?
Precipitation, Solar Radiation	**"IDAHO_EPSCOR/TERRACLIMATE"**	• Filtering to a time-range of 13.5 years (2010-2024). • Aggregation using median values.
Relative Humidity	**"NASA/GLDAS/V021/NOAH/G025/T3H"**	• Filtering to a time-range of 13.5 years (2010-2024). • Aggregation using median values.
NDMI , NDVI	**"LANDSAT/LC08/C02/T1_TOA"** • Removed Cloud Cover and Shadows from Landsat images using Qa band (Their presence can introduce noise and errors in the results.)	• Filtering to a time-range of 13.5 years (2010-2024). • Single composite image representing the median values for each pixel • .median() function is applied to the processed image collection of Landsat
Evapotranspiration	**"MODIS/006/MOD16A2"**	• Aggregation using median values.
Slope , Aspect	**NASA SRTM Digital Elevation 30m**	• **No Preprocessing** (Data and are static over short periods, no aggregation was applied.)
Land Cover	**MCD12Q1.061 MODIS Land Cover Type Yearly Global 500m**	• **No Preprocessing** (Reprojected to match the CRS of other products.)

Methodology - Data Preprocessing

Using the median to avoid the outliers (extreme values)

Imported and processed each image collection that has these factors Using Google Earth Engine here some :

Factor	Image Collection	Preprocessed ?
Population Density , Human Impact Index	Human Impact Index (HII) by Wildlife Conservation Society	• Aggregating to the **mean** values
Fuel Load	"NASA/ORNL/biomass_carbon_density/v1"	• Aggregating to the **median** values
Fuel Moisture Content	"MODIS/006/MOD13Q1"	• Aggregating to the **median** values
Lightning Frequency	"NASA/GLDAS/V021/NOAH/G025/T3H"	• Aggregating to the **median** values
		median values.
Slope , Aspect	NASA SRTM Digital Elevation 30m	• **No Preprocessing** (Data and are static over short periods, no aggregation was applied.)
Land Cover	MCD12Q1.061 MODIS Land Cover Type Yearly Global 500m	• **No Preprocessing** (Reprojected to match the CRS of other products.)

Red = Burned (1)
White = Not Burned (0)

Data Ingestion – How?

Load Continent to GEE

As mentioned before each continent is loaded as an asset to GEE in order to reduce the computation power needed.

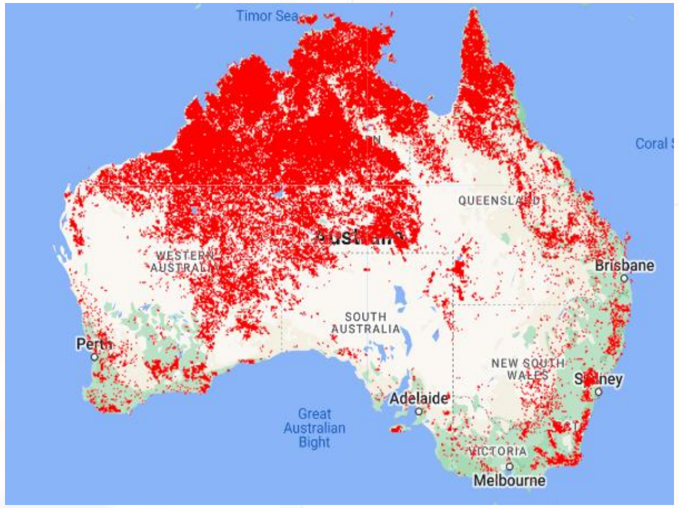
Red = Burned (1)
White = Not Burned (0)

Data Ingestion – How?

Load Continent to GEE

As mentioned before each continent is loaded as an asset to GEE in order to reduce the computation power needed.

Pure Fires binary mask



1-Load MODIS burned area collection (Filter by start and end date)
2-Define burn date threshold (Threshold = 1)
3-Classify pixels (Function to classify pixels as burned if burn date > threshold)
4-Create binary mask per each image
5-Create cumulative burned area image (Sum the binary masks from the collection)
6-Create Pure Fires binary mask

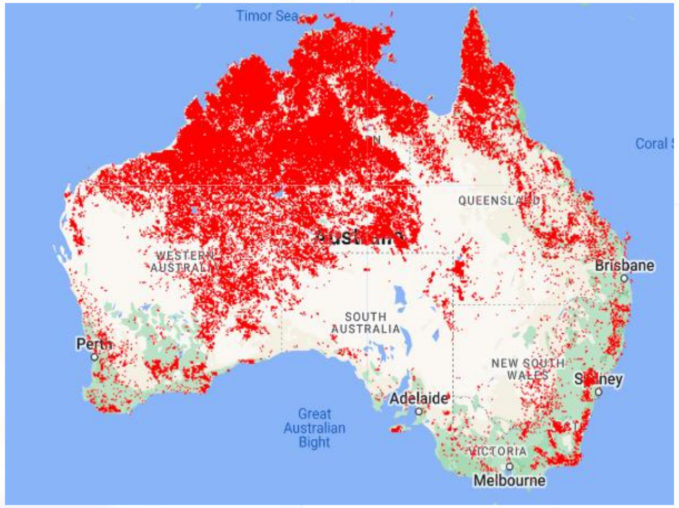
Red = Burned (1)
White = Not Burned (0)

Data Ingestion – How?

Load Continent to GEE

As mentioned before each continent is loaded as an asset to GEE in order to reduce the computation power needed.

Pure Fires binary mask



1-Load MODIS burned area collection
(Filter by start and end date)
2-Define burn date threshold
(Threshold = 1)
3-Classify pixels (Function to classify pixels as burned if burn date > threshold)
4-Create binary mask per each image
5-Create cumulative burned area image
(Sum the binary masks from the collection)
6-Create Pure Fires binary mask

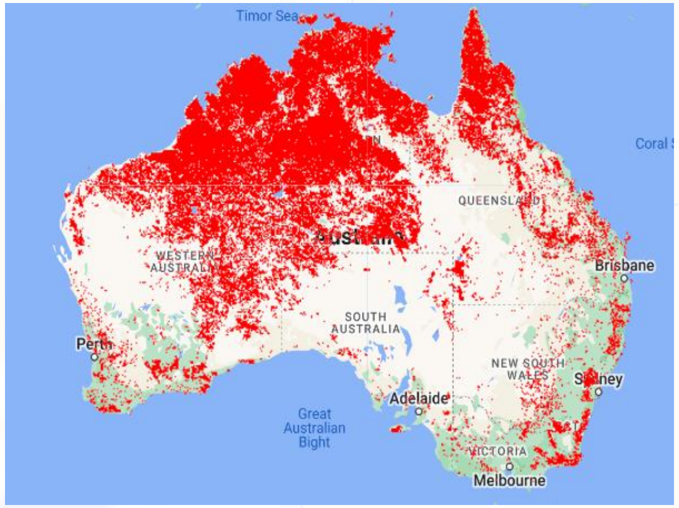
Burned/Not Hexagons

1-Create binary burned mask
Multiply by pixel area to get an area image.
2-Perform zonal statistics
Use reduceRegions with sum() reducer on the area image within each hexagon.
3-Calculate hexagons:
Divide into hexagons with burned pixel sum 0 and burned pixel sum not 0.
*** Each hexagon has its own id**

Red = Burned (1)
White = Not Burned (0)

Hexagon-based sampling – How?

Pure Fires binary mask

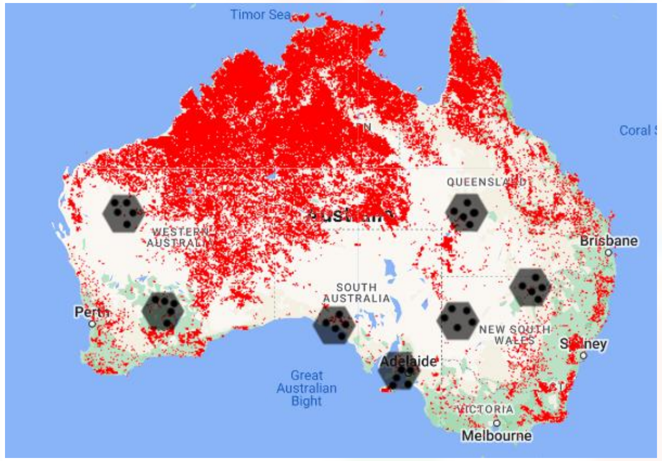


1-Load MODIS burned area collection
(Filter by start and end date)
2-Define burn date threshold
(Threshold = 1)
3-Classify pixels (Function to classify pixels as burned if burn date > threshold)
4-Create binary mask per each image
5-Create cumulative burned area image
(Sum the binary masks from the collection)
6-Create Pure Fires binary mask

Red = Burned (1)
White = Not Burned (0)

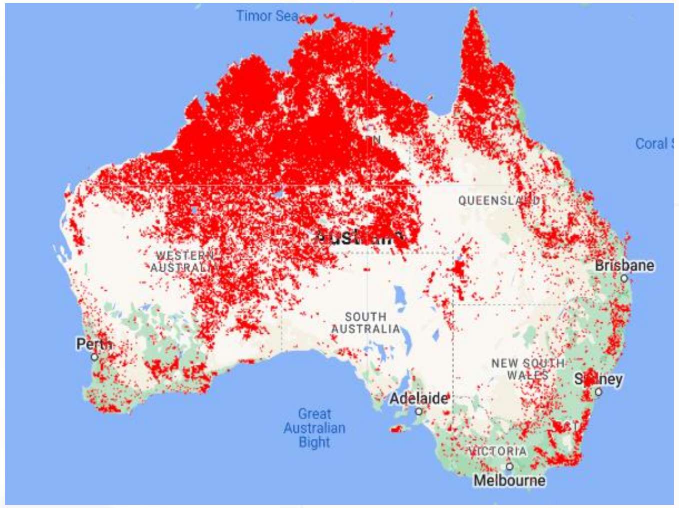
Hexagon-based sampling – How?

Not Burned samples



*

Pure Fires binary mask



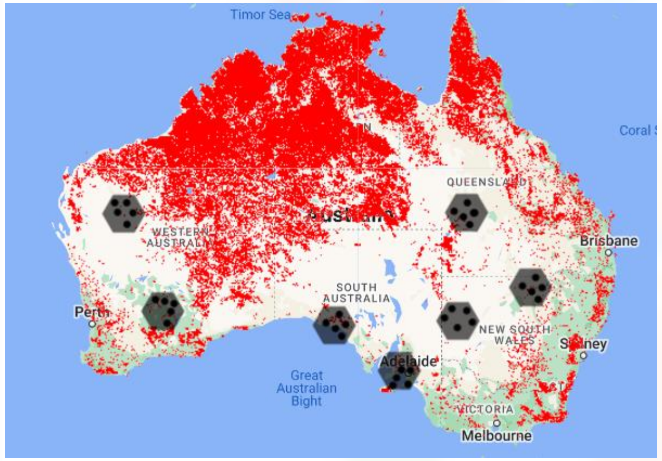
1- Create a Feature Collection:
•Represent hexagons with Not burned pixels.
2- Access each hexagon with burned pixel sum is 0:
•Extract the Not burned area information within each hexagon by id.
3- Generate a number of sample points within the Not burned areas.
4- Extract values from the predictor image variables for each sampled point.

1-Load MODIS burned area collection (Filter by start and end date)
2-Define burn date threshold (Threshold = 1)
3-Classify pixels (Function to classify pixels as burned if burn date > threshold)
4-Create binary mask per each image
5-Create cumulative burned area image (Sum the binary masks from the collection)
6-Create Pure Fires binary mask

Red = Burned (1)
White = Not Burned (0)

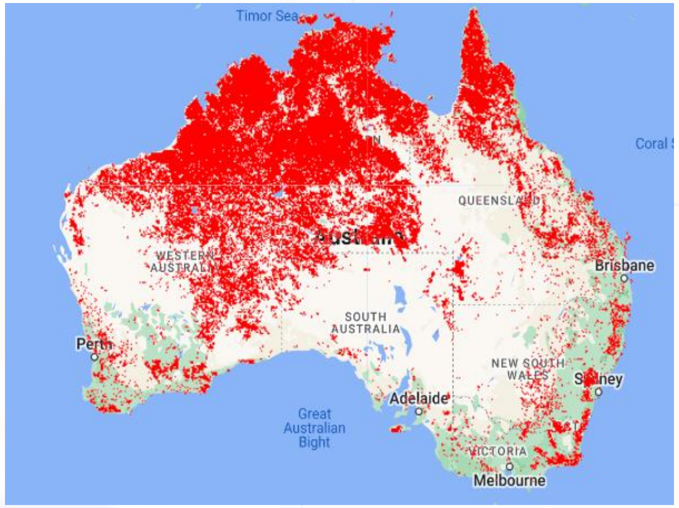
Hexagon-based sampling – How?

Not Burned samples



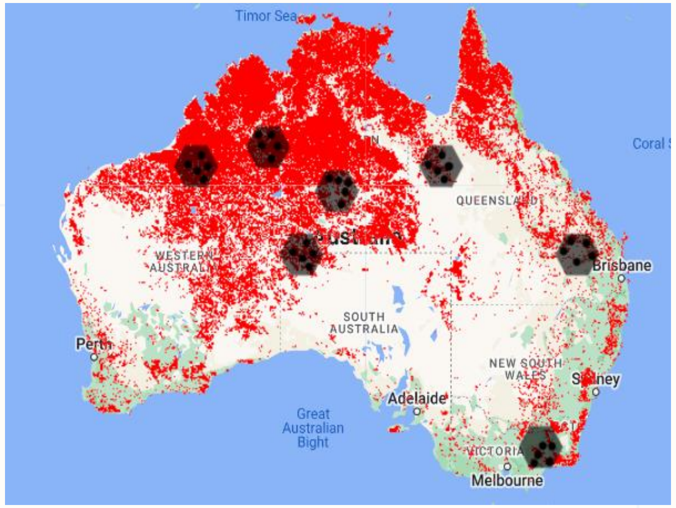
*

Pure Fires binary mask



*

Burned samples



1- Create a Feature Collection:
•Represent hexagons with Not burned pixels.
2- Access each hexagon with burned pixel sum is 0:
•Extract the Not burned area information within each hexagon by id.
3- Generate a number of sample points within the Not burned areas.
4- Extract values from the predictor image variables for each sampled point.

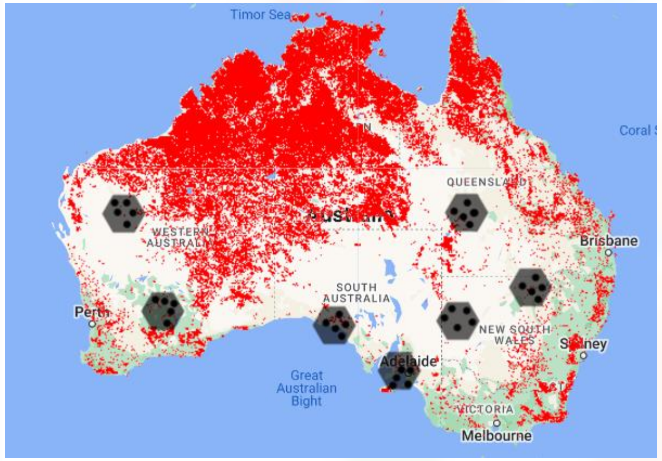
1-Load MODIS burned area collection (Filter by start and end date)
2-Define burn date threshold (Threshold = 1)
3-Classify pixels (Function to classify pixels as burned if burn date > threshold)
4-Create binary mask per each image
5-Create cumulative burned area image (Sum the binary masks from the collection)
6-Create Pure Fires binary mask

1- Create a Feature Collection:
•Represent hexagons with burned pixels.
2- Access each hexagon with burned pixel sum not 0:
•Extract the burned area information within each hexagon By id.
3- Generate a number of sample points within the burned areas.
4- Extract values from the predictor image variables for each sampled point.

Red = Burned (1)
White = Not Burned (0)

Hexagon-based sampling – How?

Not Burned samples



1- Create a Feature Collection:
•Represent hexagons with Not burned pixels.
2- Access each hexagon with burned pixel sum is 0:
•Extract the Not burned area information within each hexagon by id.
3- Generate a number of sample points within the Not burned areas.
4- Extract values from the predictor image variables for each sampled point.

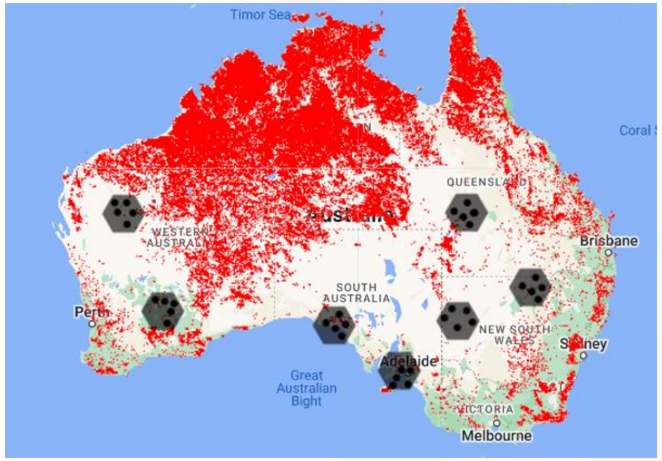
Continent	# of Hexagons	# of points within each hexagon	# of samples
Asia	37	4	148
Africa	20	5	100
North America	16	5	80
South America	12	5	60
Europe	7	7	49
Australia	7	5	35
Antarctica	7	4	28

500 sample

Red = Burned (1)
White = Not Burned (0)

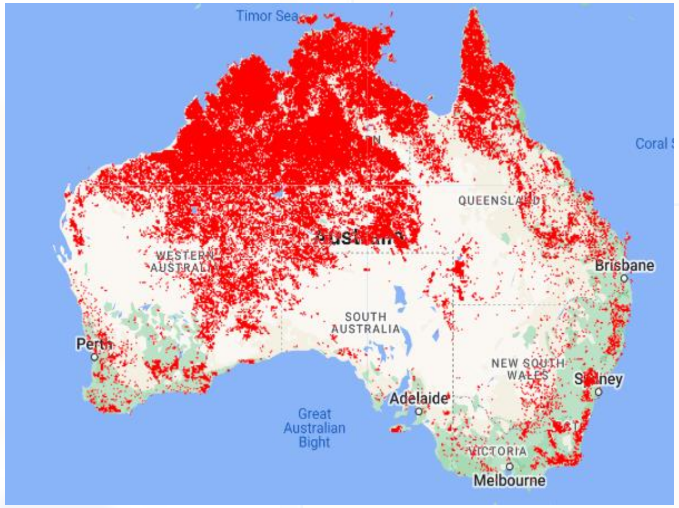
Hexagon-based sampling – How?

Not Burned samples



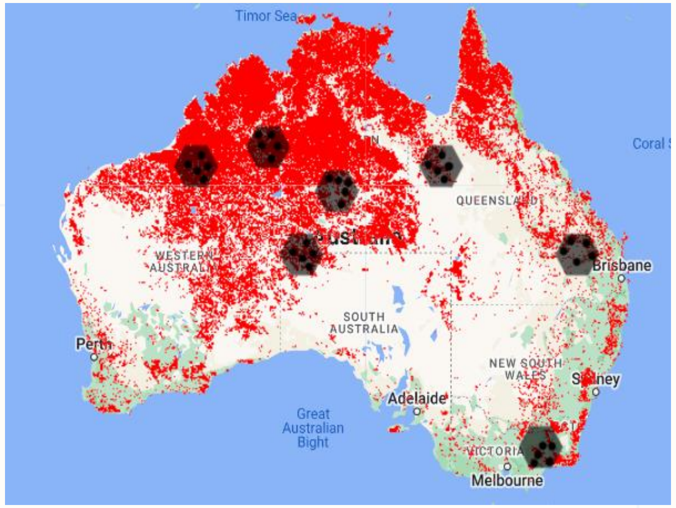
1- Create a Feature Collection:
•Represent hexagons with Not burned pixels.
2- Access each hexagon with burned pixel sum is 0:
•Extract the Not burned area information within each hexagon by id.
3- Generate a number of sample points within the Not burned areas.
4- Extract values from the predictor image variables for each sampled point.

Pure Fires binary mask



1-Load MODIS burned area collection
(Filter by start and end date)
2-Define burn date threshold
(Threshold = 1)
3-Classify pixels (Function to classify pixels as burned if burn date > threshold)
4-Create binary mask per each image
5-Create cumulative burned area image
(Sum the binary masks from the collection)
6-Create Pure Fires binary mask

Burned samples



1- Create a Feature Collection:
•Represent hexagons with burned pixels.
2- Access each hexagon with burned pixel sum not 0:
•Extract the burned area information within each hexagon By id.
3- Generate a number of sample points within the burned areas.
4- Extract values from the predictor image variables for each sampled point.

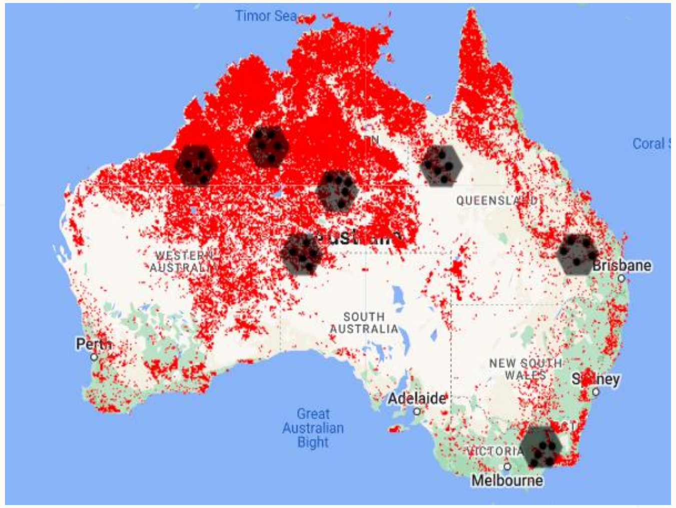
Red = Burned (1)
White = Not Burned (0)

Hexagon-based sampling – How?

Continent	# of Hexagons	# of points within each hexagon	# of samples
Asia	37	4	148
Africa	20	5	100
North America	16	5	80
South America	12	5	60
Europe	7	7	49
Australia	7	5	35
Antarctica	7	4	28

500 sample

Burned samples



1- Create a Feature Collection:
•Represent hexagons with burned pixels.
2- Access each hexagon with burned pixel sum not 0:
•Extract the burned area information within each hexagon By id.
3- Generate a number of sample points within the burned areas.
4- Extract values from the predictor image variables for each sampled point.

Model Features

After exporting the **1000 samples**, we obtained a table containing **19 factors** that will be used as features in **various models:**

Environmental (11 Factors)

- NDMI
- NDVI
- Land Cover
- Precipitation
- Elevation
- Slope
- Soil Moisture
- Temp Max
- Temp Min
- Solar Radiation
- Evapotranspiration

Human-Related (2 Factors)

- Human Impact Index Mean
- Population Density

Climate (5 Factors)

- Wind Speed
- Relative Humidity
- Lightning Frequency
- Fuel Load
- Fuel Moisture Content

Model Target

After exporting the **1000 samples**, we obtained a table containing **19 factors** that will be used as features in **various models**:

Environmental (11 Factors)

- NDMI
- NDVI
- Land Cover
- Precipitation
- Elevation
- Slope
- Soil Moisture
- Temp Max
- Temp Min
- Solar Radiation
- Evapotranspiration

Human-Related (2 Factors)

- Human Impact Index Mean
- Population Density

Climate (5 Factors)

- Wind Speed
- Relative Humidity
- Lightning Frequency
- Fuel Load
- Fuel Moisture Content

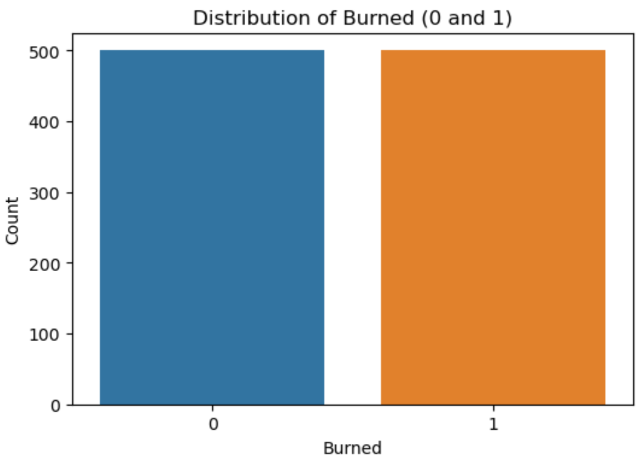
Target

- Burned (0 = Not Burned, 1 = Burned)

***We will predict the target using the provided features to determine which features are the most important.**

Building The Models

Step 0 :Exploratory Data Analysis

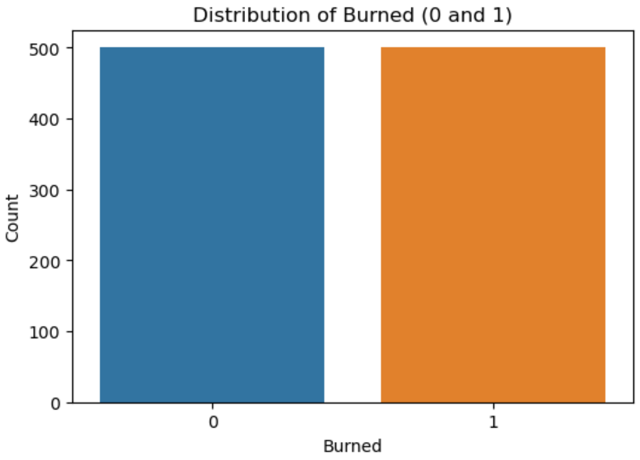


Our data is well balanced with 500 burned and 500 not burned.

Other EDA steps such as correlation matrix and histograms in the Notebook.

Building The Models

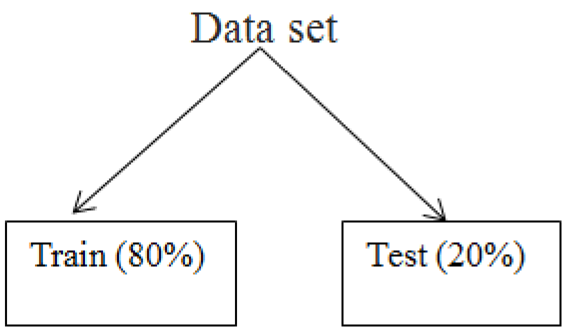
Step 0 :Exploratory Data Analysis



Our data is well balanced with 500 burned and 500 not burned.

Other EDA steps such as correlation matrix and histograms in the Notebook.

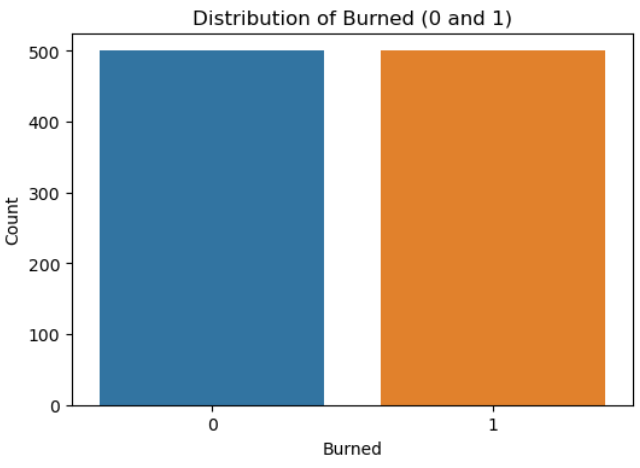
Step 1: Data Preparation



- **Data Splitting:** Split the data into training and testing sets.
  (Training set : 80% Testing set: 20%)
- **Feature Scaling:** Standardize or normalize the features to ensure that they have similar scales.

Building The Models

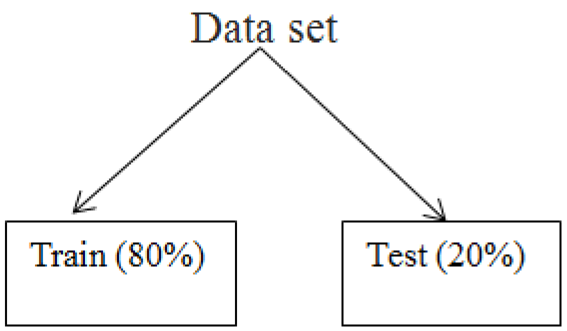
Step 0 :Exploratory Data Analysis



Our data is well balanced with 500 burned and 500 not burned.

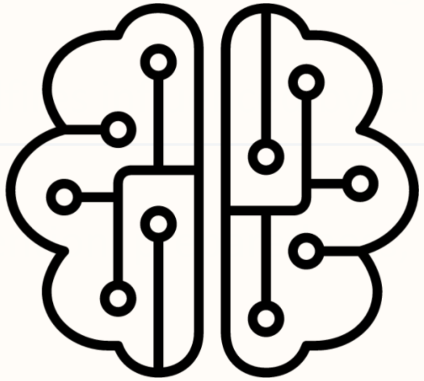
Other EDA steps such as correlation matrix and histograms in the Notebook.

Step 1: Data Preparation



- **Data Splitting:** Split the data into training and testing sets.
  (Training set : 80% Testing set: 20%)
- **Feature Scaling:** Standardize or normalize the features to ensure that they have similar scales.

Step 2: Model Selection



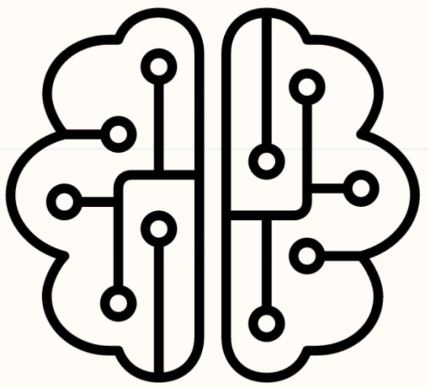
This is a **binary classification** problem so we will choose :

- **Logistic Regression**
- **Random Forest Classifier**
- **Gradient Boosting Classifier**
- **Support Vector Machine (SVM)**
- **XGBoost Classifier**

+Neural Network

Building The Models

Step 2: Model Selection



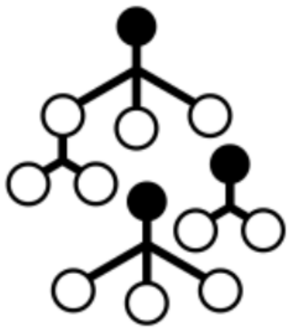
This is a **binary classification** problem so we will choose :

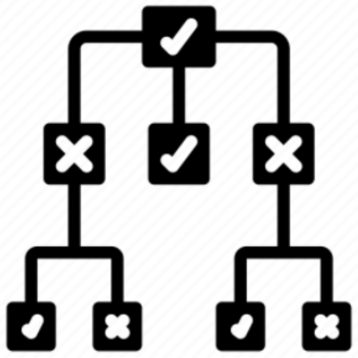
- **Logistic Regression**
- **Random Forest Classifier**
- **Gradient Boosting Classifier**
- **Support Vector Machine (SVM)**
- **XGBoost Classifier**

+Neural Network

Building The Models

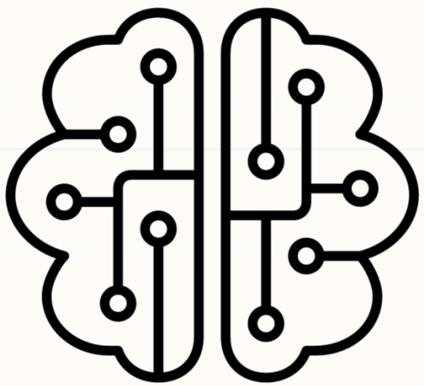
Step 3: Training the Model





- **Fit the selected models on the training data.**

Step 2: Model Selection



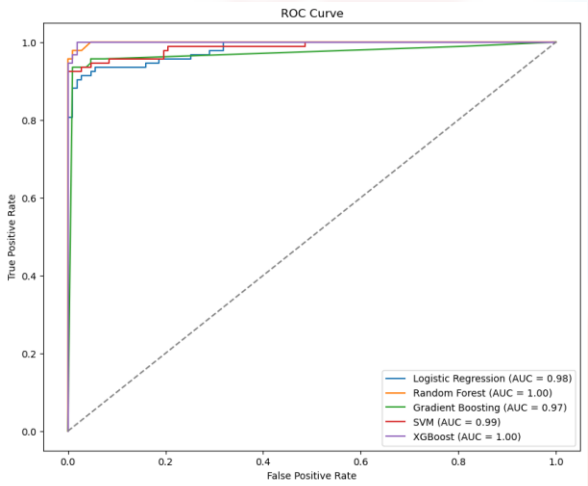
This is a **binary classification** problem so we will choose :

- **Logistic Regression**
- **Random Forest Classifier**
- **Gradient Boosting Classifier**
- **Support Vector Machine (SVM)**
- **XGBoost Classifier**

+Neural Network

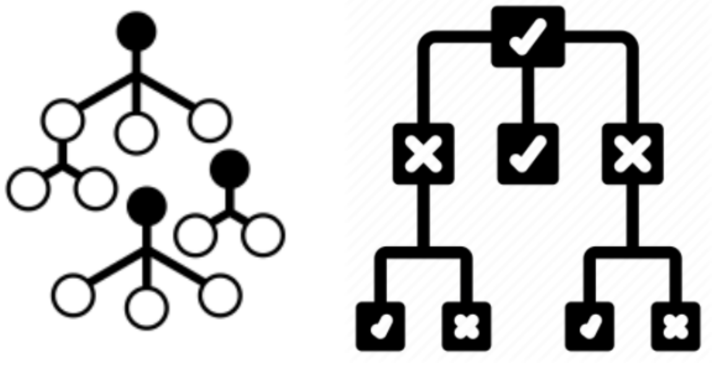
Building The Models

Step 4: Model Evaluation



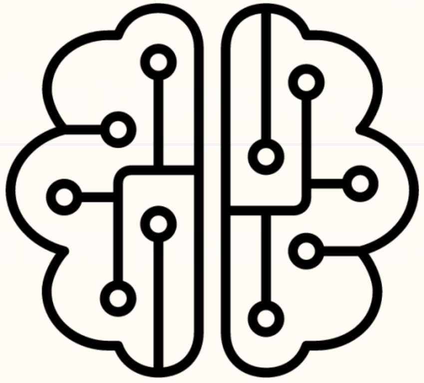
- **Predictions:** Use the trained models to make predictions on the test data.
- **Performance Metrics:** Evaluate the performance of the models using appropriate metrics such as accuracy, precision, recall, F1 score, ROC-AUC, etc.

Step 3: Training the Model



- **Fit the selected models on the training data.**

Step 2: Model Selection



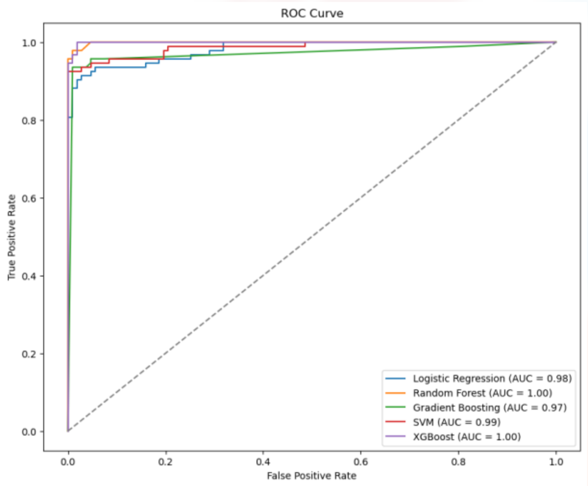
This is a **binary classification** problem so we will choose :

- **Logistic Regression**
- **Random Forest Classifier**
- **Gradient Boosting Classifier**
- **Support Vector Machine (SVM)**
- **XGBoost Classifier**

+Neural Network

Building The Models

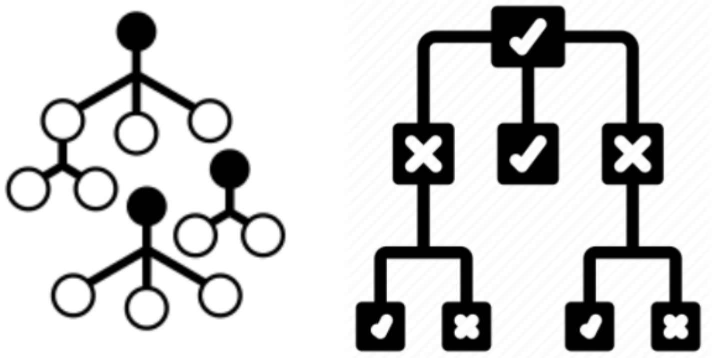
Step 4: Model Evaluation



- **Predictions:** Use the trained models to make predictions on the test data.

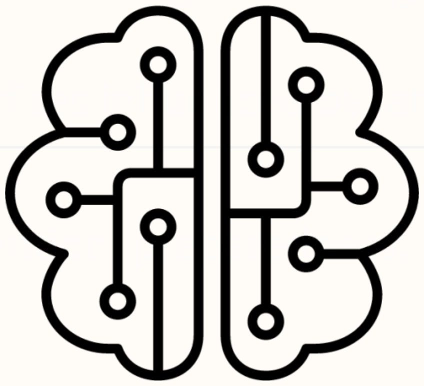
- **Performance Metrics:** Evaluate the performance of the models using appropriate metrics such as accuracy, precision, recall, F1 score, ROC-AUC, etc.

Step 3: Training the Model



- **Fit the selected models on the training data.**

Step 2: Model Selection



This is a **binary classification** problem so we will choose :

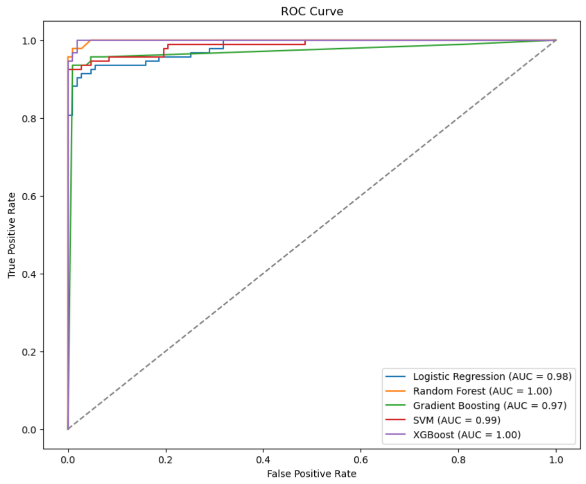
- **Logistic Regression**
- **Random Forest Classifier**
- **Gradient Boosting Classifier**
- **Support Vector Machine (SVM)**
- **XGBoost Classifier**

+Neural Network

Step 5 : Find the most important features that affecting wildfire distribution using the best 4 models.

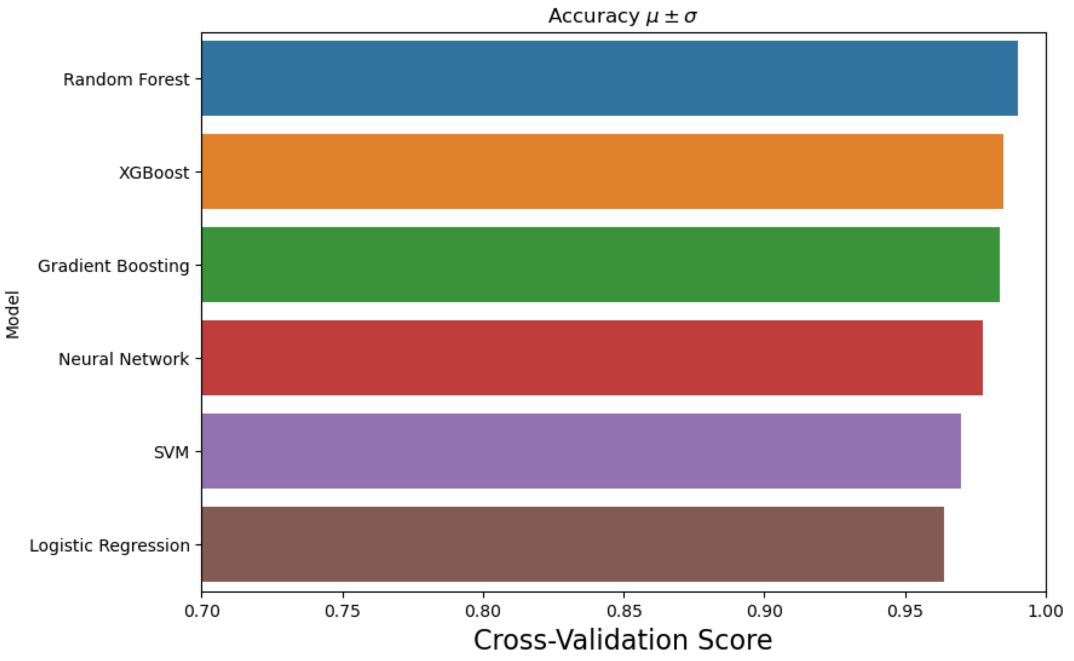
Evaluating The Models

Step 4: Model Evaluation



- **Predictions:** Use the trained models to make predictions on the test data.
- **Performance Metrics:** Evaluate the performance of the models using appropriate metrics such as accuracy, precision, recall, F1 score, ROC-AUC, etc.

Step 4.a: Cross-Validation

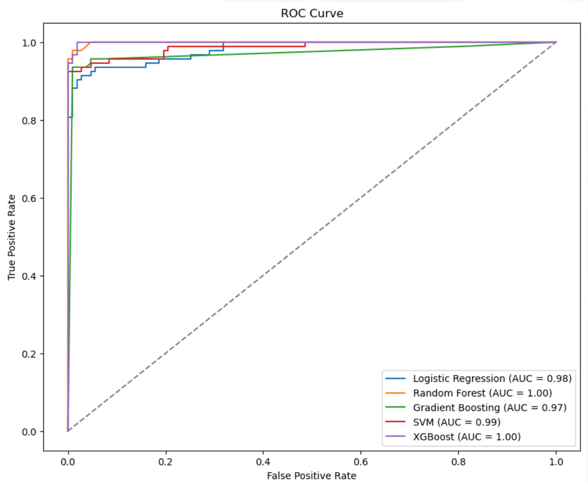


- **Cross-validation** involves splitting the training data into multiple folds (subsets), training the model on some folds while validating it on the remaining fold, and repeating this process multiple times.

***The mean cross-validation** score is very close to the training accuracy (98.9%), and the **standard deviation is relatively low**, indicating that **the model is performing consistently across different folds and is not overfitting.**

Evaluating The Models

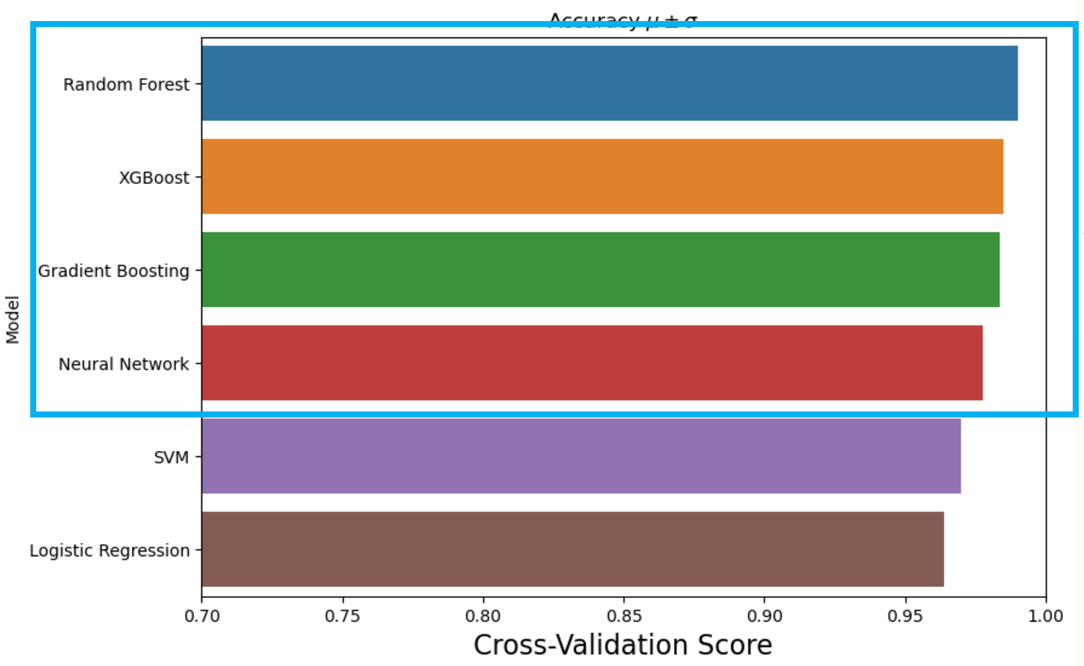
Step 4: Model Evaluation



- **Predictions:** Use the trained models to make predictions on the test data.

- **Performance Metrics:** Evaluate the performance of the models using appropriate metrics such as accuracy, precision, recall, F1 score, ROC-AUC, etc.

Step 4.a: Cross-Validation

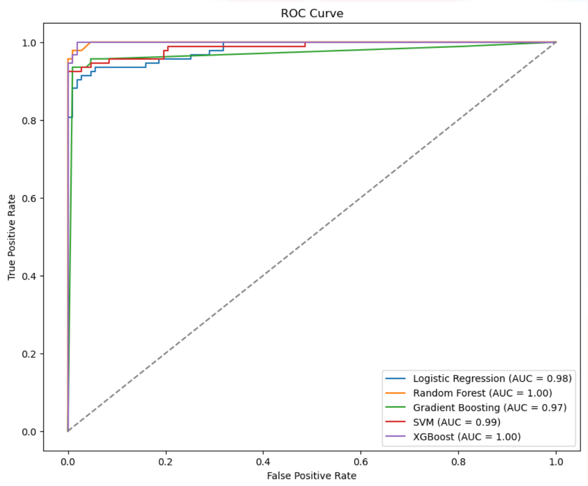


We will continue with top 4 models

- **Cross-validation** involves splitting the training data into multiple folds (subsets), training the model on some folds while validating it on the remaining fold, and repeating this process multiple times.

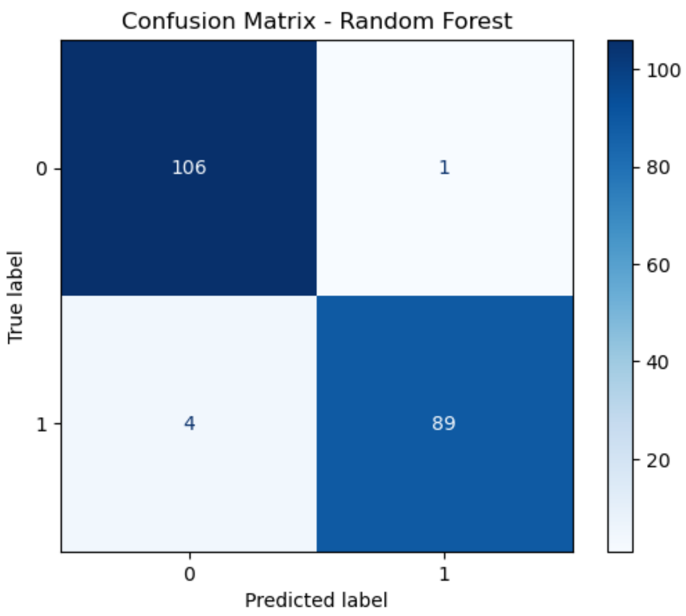
Evaluating The Models

Step 4: Model Evaluation

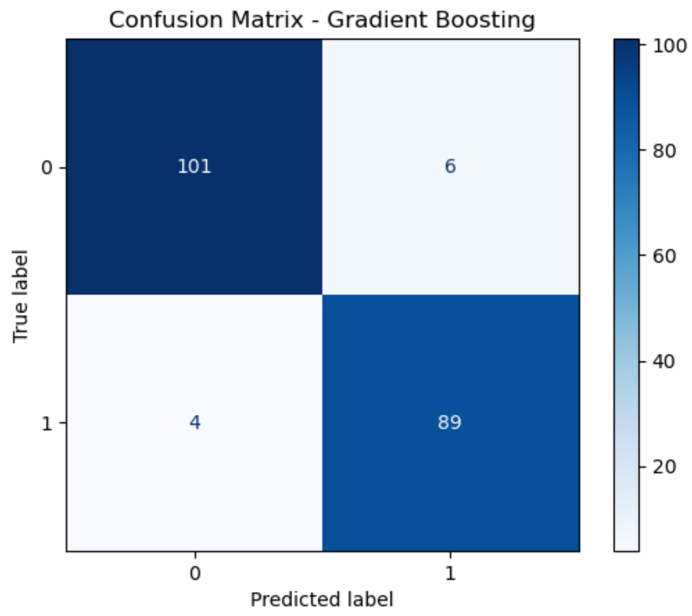


- **Predictions:** Use the trained models to make predictions on the test data.
- **Performance Metrics:** Evaluate the performance of the models using appropriate metrics such as accuracy, precision, recall, F1 score, ROC-AUC, etc.

Step 4.b: Confusion Matrix



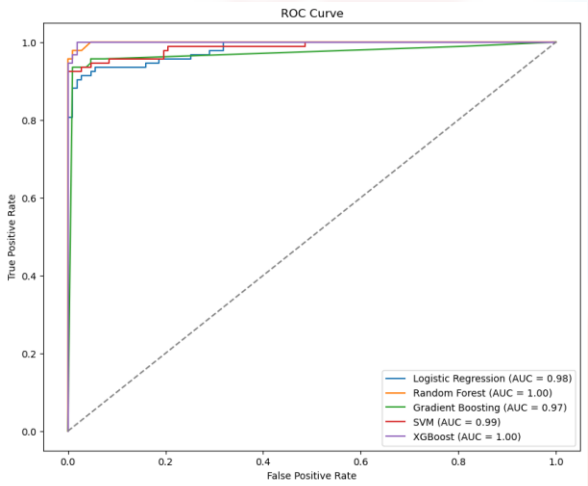

```
Model: Random Forest
Accuracy: 0.9750
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.99      0.98       107
           1       0.99      0.96      0.97        93

    accuracy                           0.97       200
   macro avg       0.98      0.97      0.97       200
weighted avg       0.98      0.97      0.97       200
```




```
Model: Gradient Boosting
Accuracy: 0.9500
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.94      0.95       107
           1       0.94      0.96      0.95        93

    accuracy                           0.95       200
   macro avg       0.95      0.95      0.95       200
weighted avg       0.95      0.95      0.95       200
```

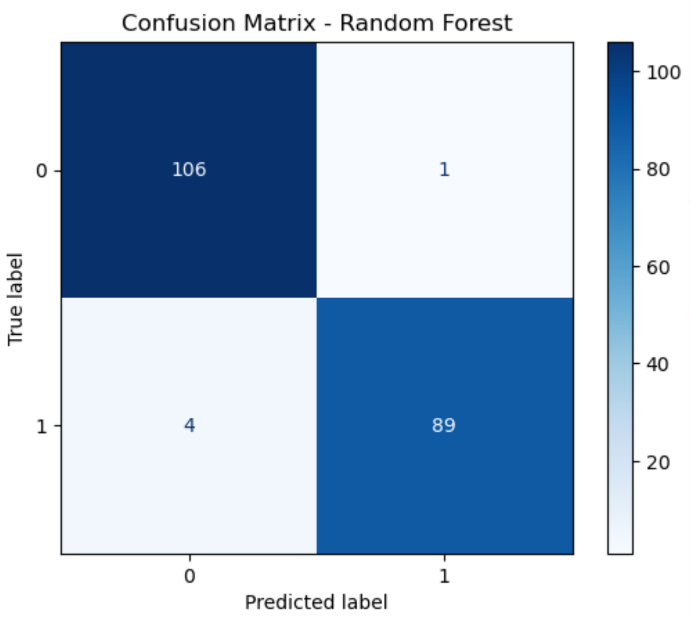
Evaluating The Models

Step 4: Model Evaluation



- **Predictions:** Use the trained models to make predictions on the test data.
- **Performance Metrics:** Evaluate the performance of the models using appropriate metrics such as accuracy, precision, recall, F1 score, ROC-AUC, etc.

Step 4.b: Confusion Matrix

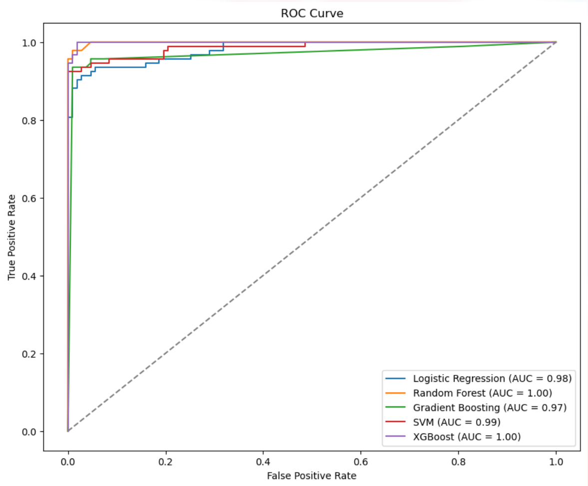



```
Model: Random Forest
Accuracy: 0.9750
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.99      0.98       107
           1       0.99      0.96      0.97        93

    accuracy                           0.97       200
   macro avg       0.98      0.97      0.97       200
weighted avg       0.98      0.97      0.97       200
```

- These metrics collectively indicate that the **Random Forest model is highly effective in distinguishing between the two classes with minimal errors.**
- The model's **high precision** ensures that it makes very few incorrect positive predictions.
- The model's **high recall** ensures that it correctly identifies almost all actual instances of each class.
- **The f1-score,** balancing both precision and recall, further confirms the model's **robustness and reliability in classification tasks.**

Evaluating The Models

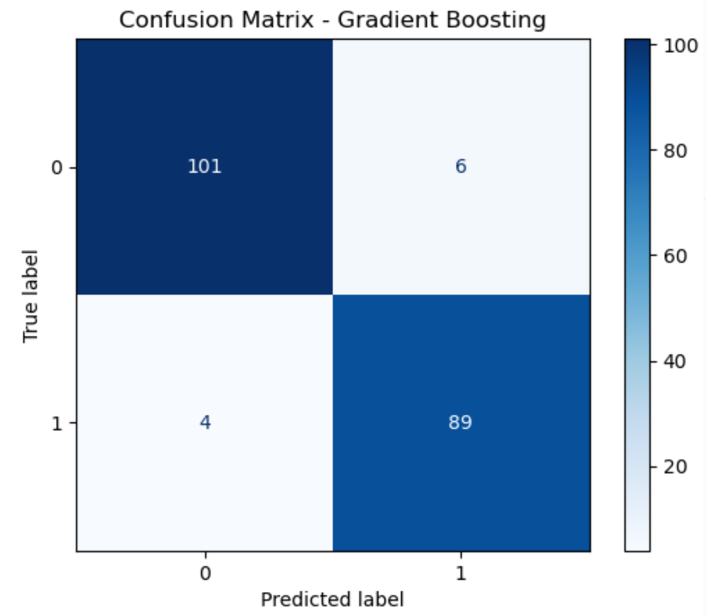
Step 4: Model Evaluation



- **Predictions:** Use the trained models to make predictions on the test data.
- **Performance Metrics:** Evaluate the performance of the models using appropriate metrics such as accuracy, precision, recall, F1 score, ROC-AUC, etc.

Step 4.b: Confusion Matrix

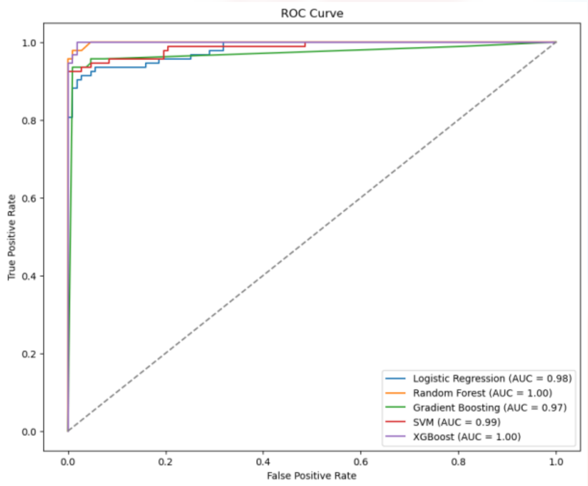
- These metrics collectively indicate that the **Gradient Boosting model is highly effective in distinguishing between the two classes with minimal errors.**
- The model's **high precision** ensures that it makes very few incorrect positive predictions.
- The model's **high recall** ensures that it correctly identifies almost all actual instances of each class.
- **The f1-score,** balancing both precision and recall, further confirms the model's robustness and reliability in classification tasks.




```
Model: Gradient Boosting
Accuracy: 0.9500
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.94      0.95       107
           1       0.94      0.96      0.95        93

    accuracy                           0.95       200
   macro avg       0.95      0.95      0.95       200
weighted avg       0.95      0.95      0.95       200
```

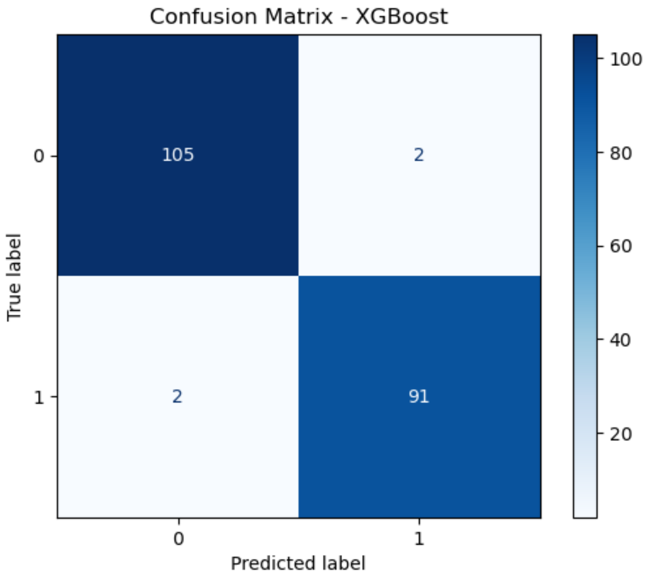
Evaluating The Models

Step 4: Model Evaluation

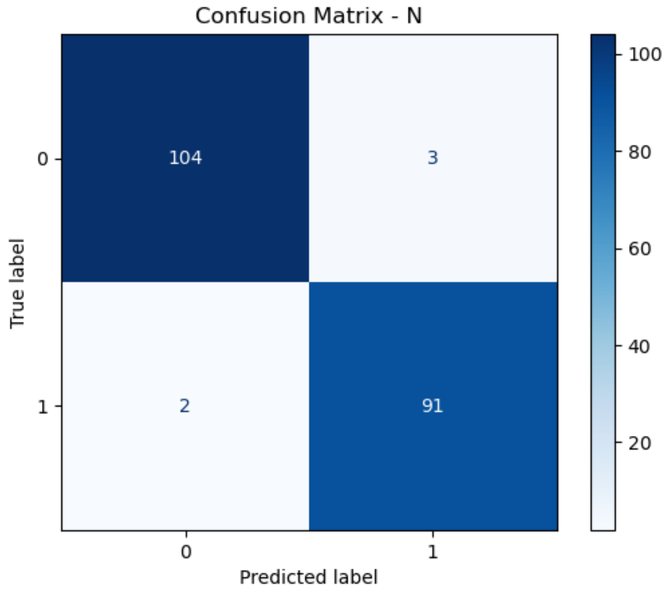


- **Predictions:** Use the trained models to make predictions on the test data.
- **Performance Metrics:** Evaluate the performance of the models using appropriate metrics such as accuracy, precision, recall, F1 score, ROC-AUC, etc.

Step 4.b: Confusion Matrix




```
Model: XGBoost
Accuracy: 0.9800
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.98      0.98       107
           1       0.98      0.98      0.98        93

    accuracy                           0.98       200
   macro avg       0.98      0.98      0.98       200
weighted avg       0.98      0.98      0.98       200
```

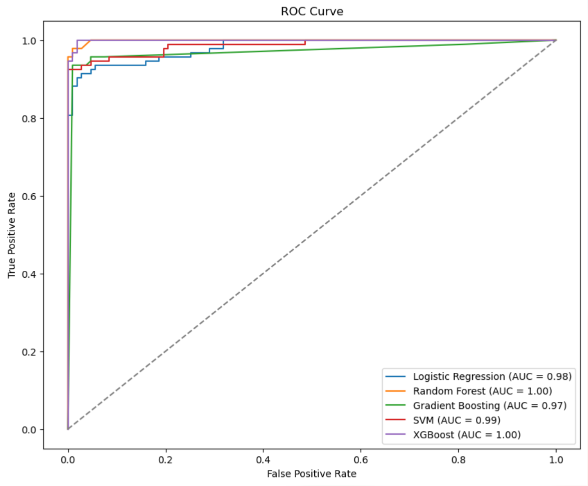



```
Model: N
Accuracy: 0.9750
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.97      0.98       107
           1       0.97      0.98      0.97        93

    accuracy                           0.97       200
   macro avg       0.97      0.98      0.97       200
weighted avg       0.98      0.97      0.98       200
```

Evaluating The Models

- **TPR is also known as sensitivity or recall.**
- **TPR = TP / (TP + FN)**
- **FPR = FP / (FP + TN)**

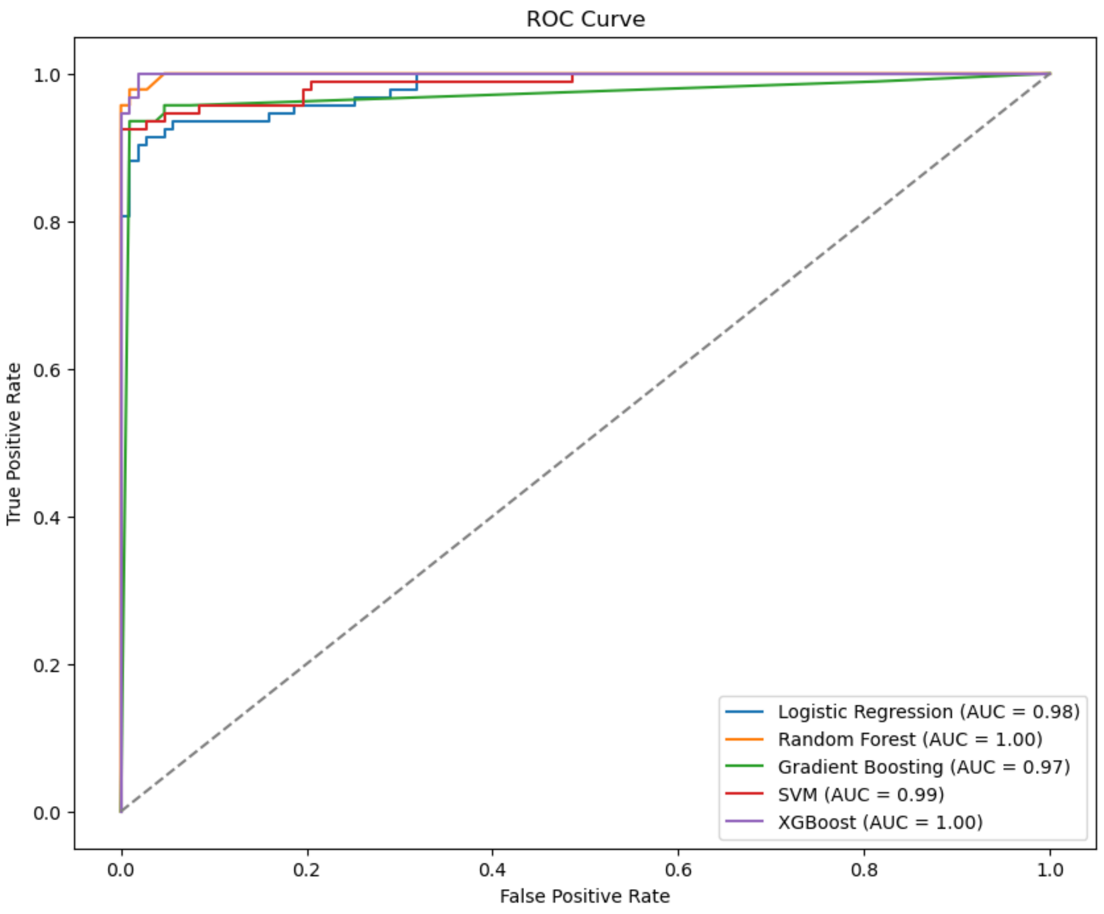
Step 4: Model Evaluation



- **Predictions:** Use the trained models to make predictions on the test data.

- **Performance Metrics:** Evaluate the performance of the models using appropriate metrics such as accuracy, precision, recall, F1 score, ROC-AUC, etc.

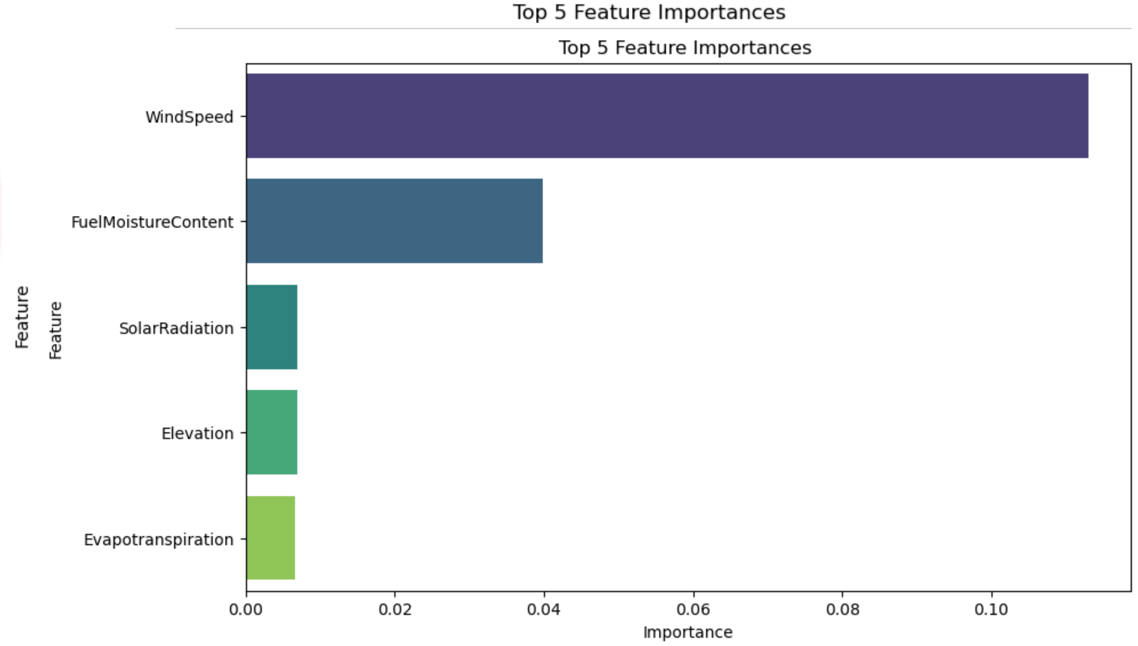
Step 4.c: ROC (Receiver Operating Characteristic) curve



- **The ROC** curve helps visualize the trade-off between the TPR and FPR for different threshold values.
- **AUC (Area Under the Curve**) summarizes the overall performance of the model: the closer the AUC is to 1, the better the model's performance.
- **All the models perform significantly well,** with Random Forest and XGBoost achieving perfect AUC scores of 1.00.

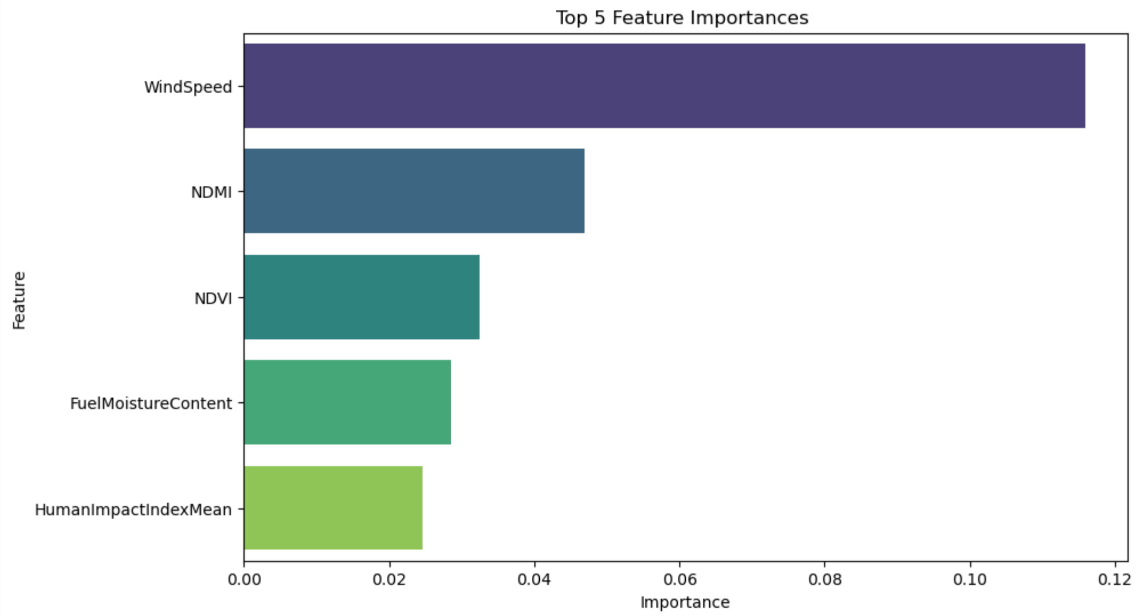
Most Important Features

Random Forest



Factor	Importance
WindSpeed	0.113000
FuelMoistureContent	0.039833
SolarRadiation	0.007000
Elevation	0.007000
Evapotranspiration	0.006667

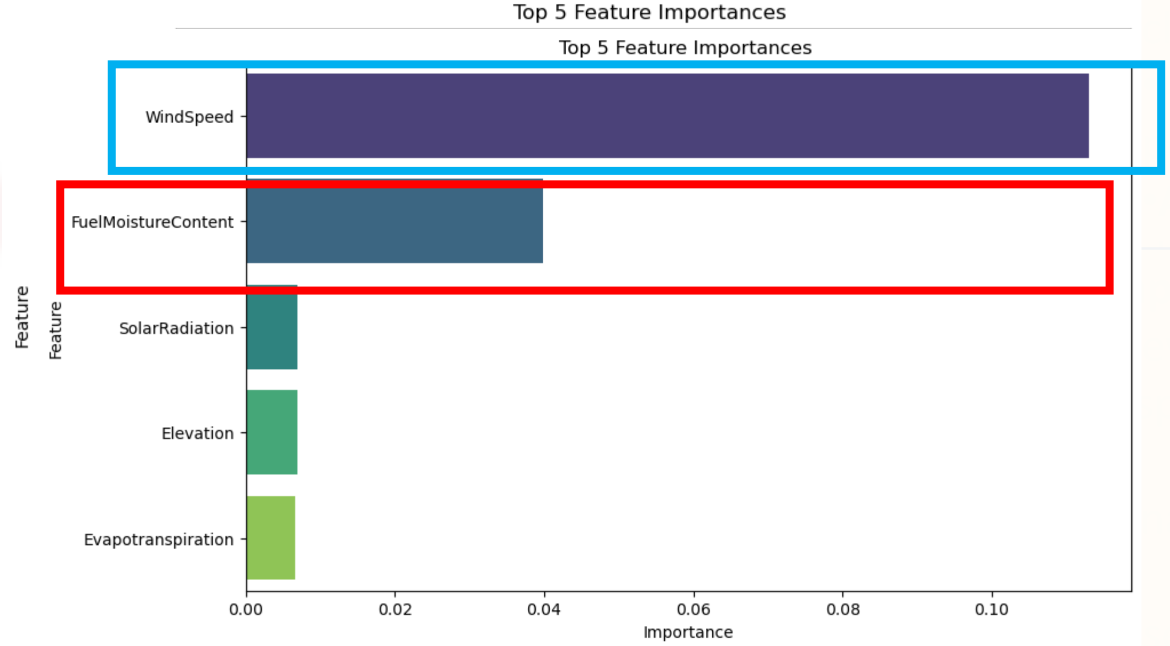
Neural Network



Feature	Importance
WindSpeed	0.116000
NDMI	0.047000
NDVI	0.032500
FuelMoistureContent	0.028500
HumanImpactIndexMean	0.024667

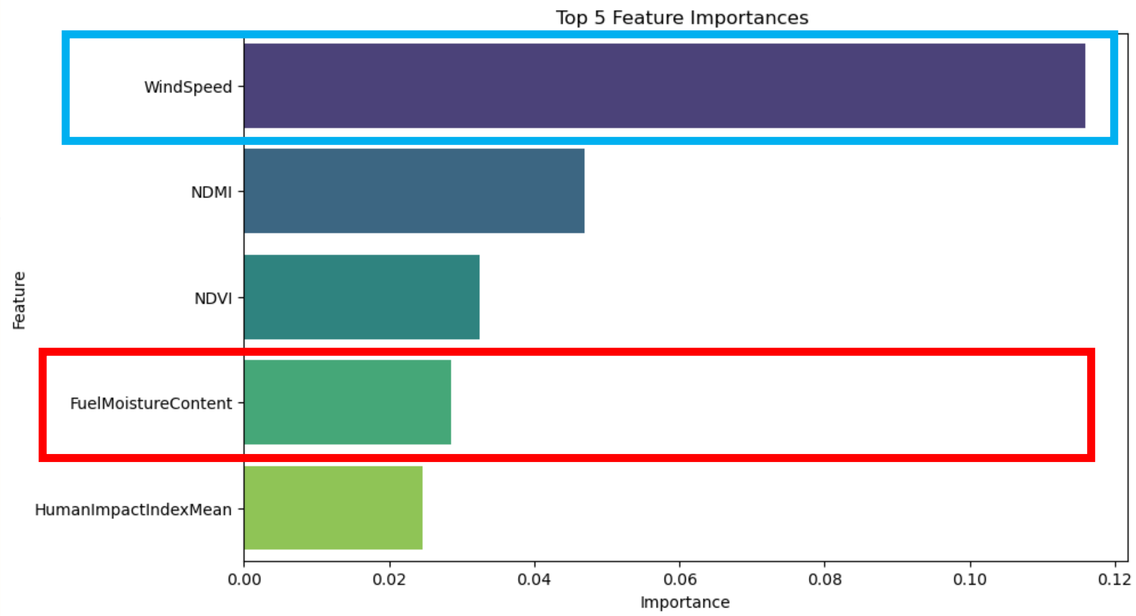
Most Important Features

Random Forest



Factor	Importance
WindSpeed	0.113000
FuelMoistureContent	0.039833
SolarRadiation	0.007000
Elevation	0.007000
Evapotranspiration	0.006667

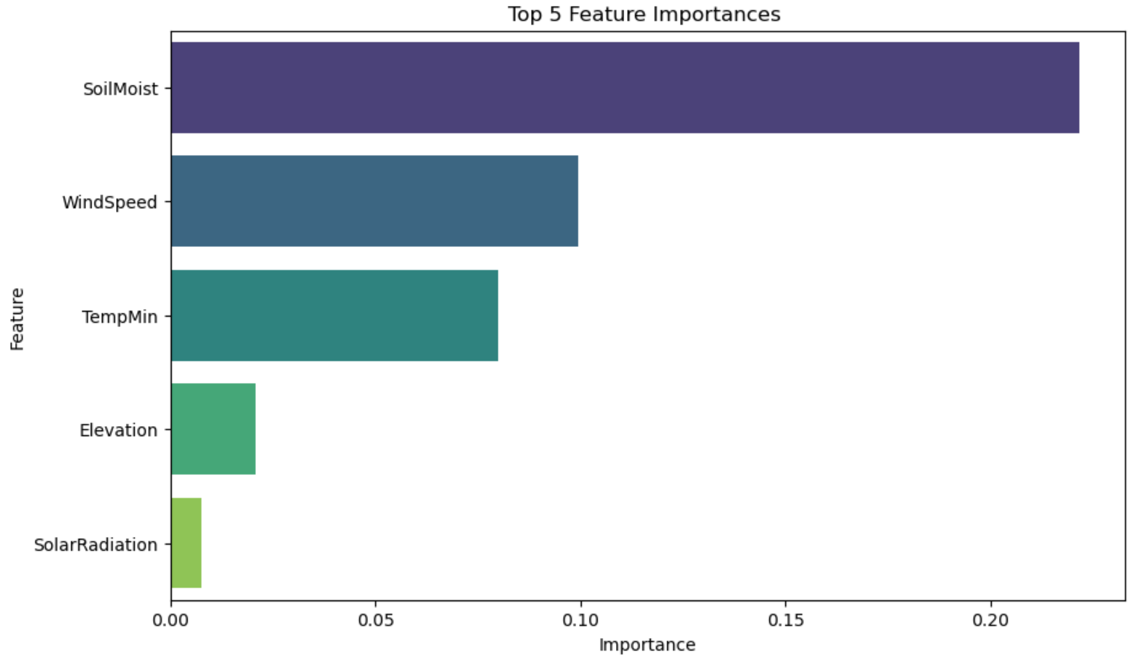
Neural Network



Feature	Importance
WindSpeed	0.116000
NDMI	0.047000
NDVI	0.032500
FuelMoistureContent	0.028500
HumanImpactIndexMean	0.024667

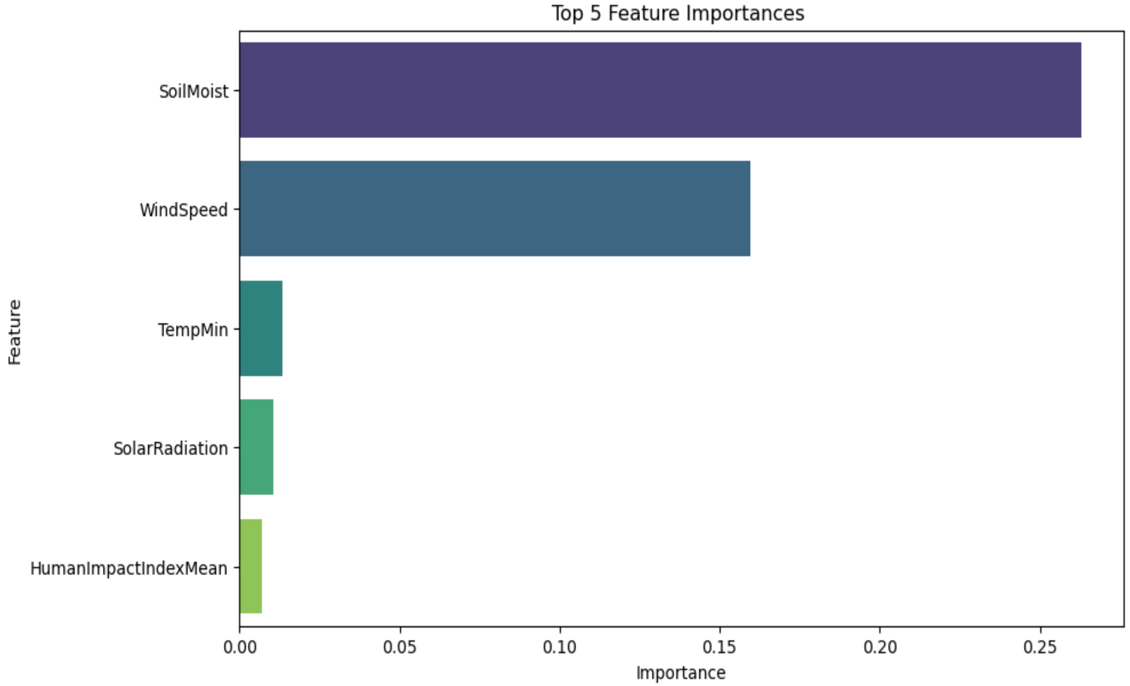
Most Important Features

Gradient boosting



Feature	Importance
SoilMoist	0.221667
WindSpeed	0.099333
TempMin	0.079833
Elevation	0.020833
SolarRadiation	0.007667

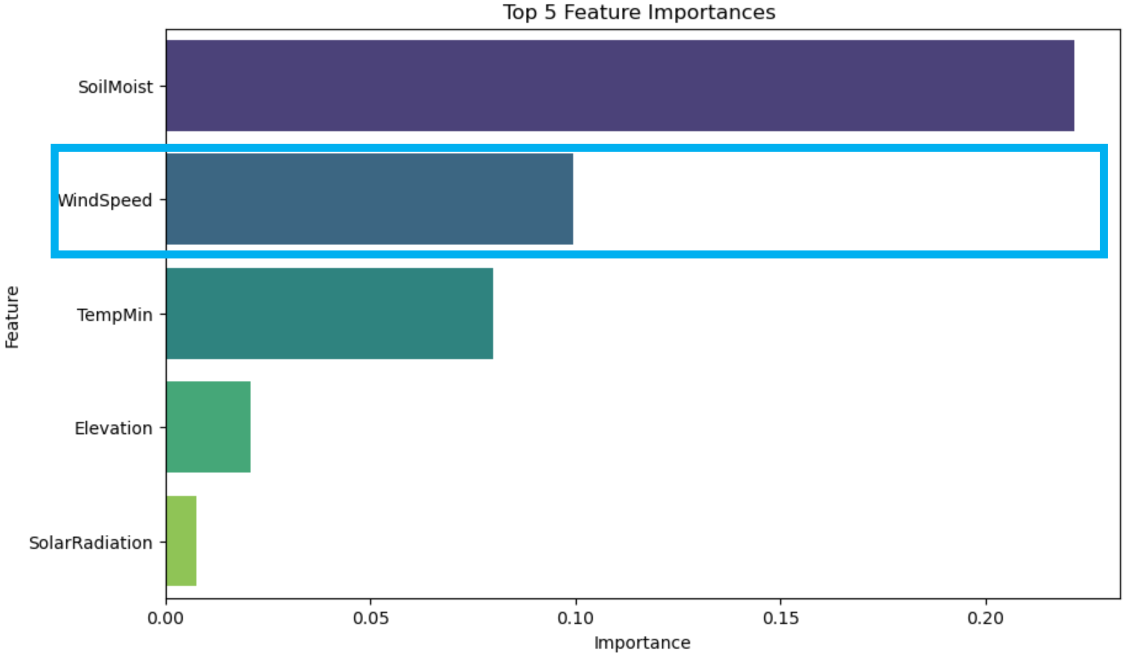
Xgboost



Feature	Importance
SoilMoist	0.263167
WindSpeed	0.159500
TempMin	0.013500
SolarRadiation	0.010500
HumanImpactIndexMean	0.006833

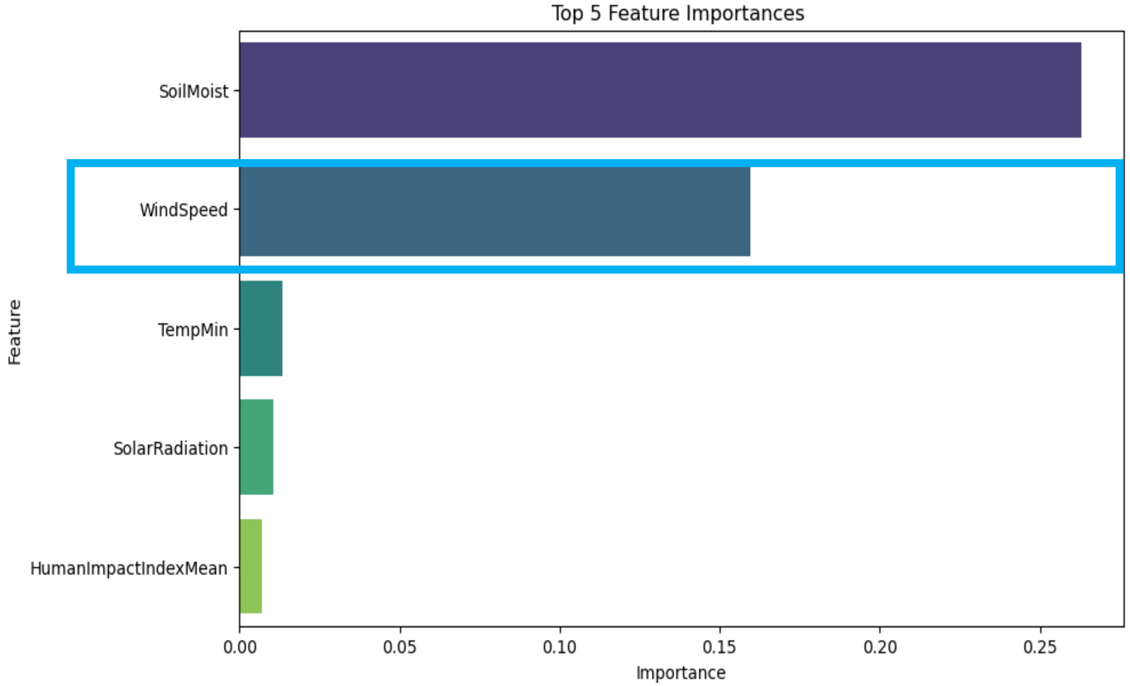
Most Important Features

Gradient boosting



Feature	Importance
SoilMoist	0.221667
WindSpeed	0.099333
TempMin	0.079833
Elevation	0.020833
SolarRadiation	0.007667

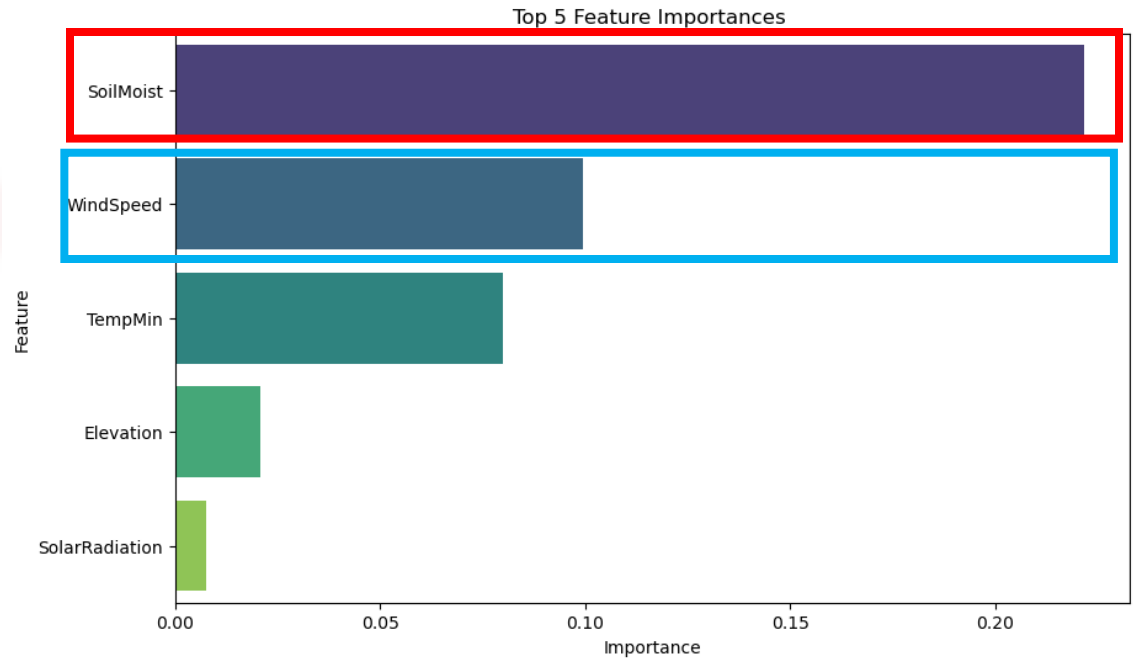
Xgboost



Feature	Importance
SoilMoist	0.263167
WindSpeed	0.159500
TempMin	0.013500
SolarRadiation	0.010500
HumanImpactIndexMean	0.006833

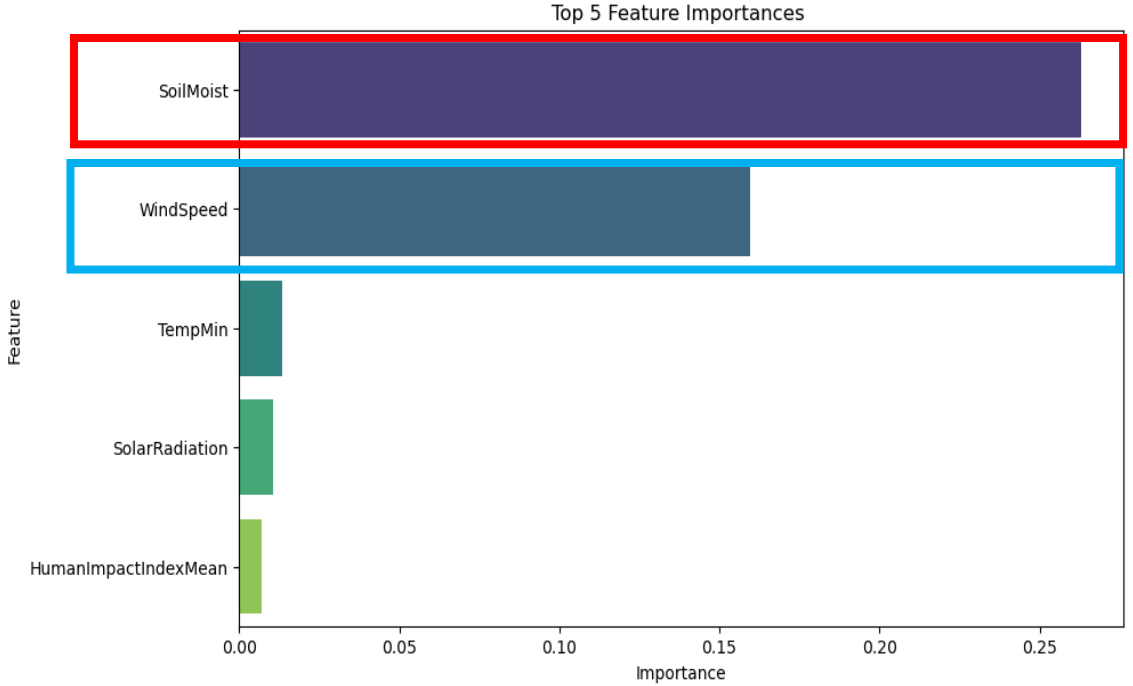
Most Important Features

Gradient boosting



Feature	Importance
SoilMoist	0.221667
WindSpeed	0.099333
TempMin	0.079833
Elevation	0.020833
SolarRadiation	0.007667

Xgboost



Feature	Importance
SoilMoist	0.263167
WindSpeed	0.159500
TempMin	0.013500
SolarRadiation	0.010500
HumanImpactIndexMean	0.006833

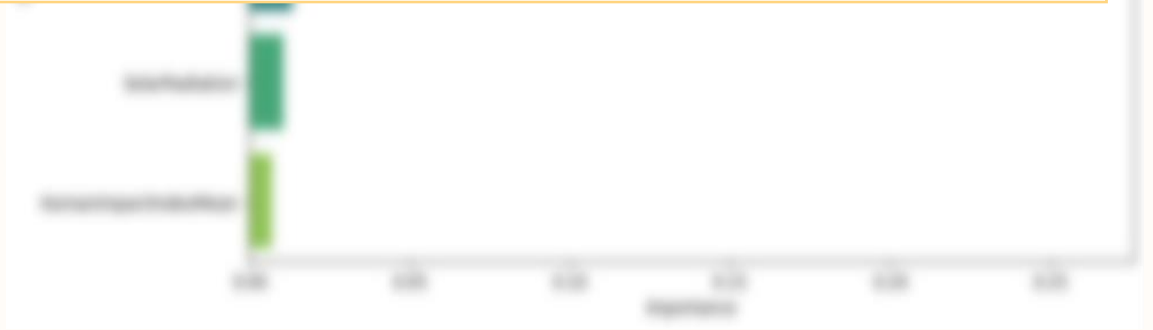
Most Important Features

All of the models agree on that **1. wind speed** is the **most crucial factor** affecting wildfire distribution. This is understandable because wind speed:

•**Increases fire spread and intensity.**

•**Enhances spotting and changes fire direction.**

•**Promotes drying of fuels, increasing fire risk.**



Most Important Features

All of the models agree on that **1. wind speed** is the **most crucial factor** affecting wildfire distribution. This is understandable because wind speed:

- **Increases fire spread and intensity.**
- **Enhances spotting and changes fire direction.**
- **Promotes drying of fuels, increasing fire risk.**

2. Soil moisture

- Soil moisture levels influence vegetation flammability and wildfire intensity.
- **Dry soil conditions extend fire seasons and increase fire frequency and spread**.

Most Important Features

All of the models agree on that **1. wind speed** is the **most crucial factor** affecting wildfire distribution. This is understandable because wind speed:

- **Increases fire spread and intensity.**
- **Enhances spotting and changes fire direction.**
- **Promotes drying of fuels, increasing fire risk.**

2. Soil moisture	**3. Solar radiation**
• Soil moisture levels influence vegetation flammability and wildfire intensity. • **Dry soil conditions extend fire seasons and increase fire frequency and spread**.	• **dries out vegetation,** making it more susceptible to burning. • Higher radiation levels can lead to more intense fires by increasing fuel flammability and combustion rates.

Most Important Features

All of the models agree on that **1. wind speed** is the **most crucial factor** affecting wildfire distribution. This is understandable because wind speed:

- **Increases fire spread and intensity.**
- **Enhances spotting and changes fire direction.**
- **Promotes drying of fuels, increasing fire risk.**

2. Soil moisture

- Soil moisture levels influence vegetation flammability and wildfire intensity.
- **Dry soil conditions extend fire seasons and increase fire frequency and spread**.

3. Solar radiation

- **dries out vegetation**, making it more susceptible to burning.
- Higher radiation levels can lead to more intense fires by increasing fuel flammability and combustion rates.

4. Elevation

- Elevated areas may experience drier conditions and stronger winds, facilitating fire spread.

Most Important Features

All of the models agree on that **1. wind speed** is the **most crucial factor** affecting wildfire distribution. This is understandable because wind speed:

- **Increases fire spread and intensity.**
- **Enhances spotting and changes fire direction.**
- **Promotes drying of fuels, increasing fire risk.**

2. Soil moisture

- Soil moisture levels influence vegetation flammability and wildfire intensity.
- **Dry soil conditions extend fire seasons and increase fire frequency and spread**.

3. Solar radiation

- **dries out vegetation,** making it more susceptible to burning.
- Higher radiation levels can lead to more intense fires by increasing fuel flammability and combustion rates.

4. Elevation

- Elevated areas may experience drier conditions and stronger winds, facilitating fire spread.

5. Aspect

- Affects how much sunlight slopes receive, influencing vegetation dryness and fire risk.

Conclusion

Key Findings:

Several models are trained (but the highest 4 are used):

- Logistic Regression
- **Random Forest Classifier**
- **Gradient Boosting Classifier**
- Support Vector Machine (SVM)
- **XGBoost Classifier**
- **+Neural Network**

Top 5 Important Features are:

- Wind speed
- Soil Moisture
- Solar Radiation
- Elevation
- Aspect

Developed upon an existing method:

- **Adjusting grid size** simplifies computations and reduces computational strain.
- **Introducing additional features** increases model variability.
- Modifying the number of hexagons simplifies looping processes.
- Making it UpToDate (2010-Now)

One More Thing..

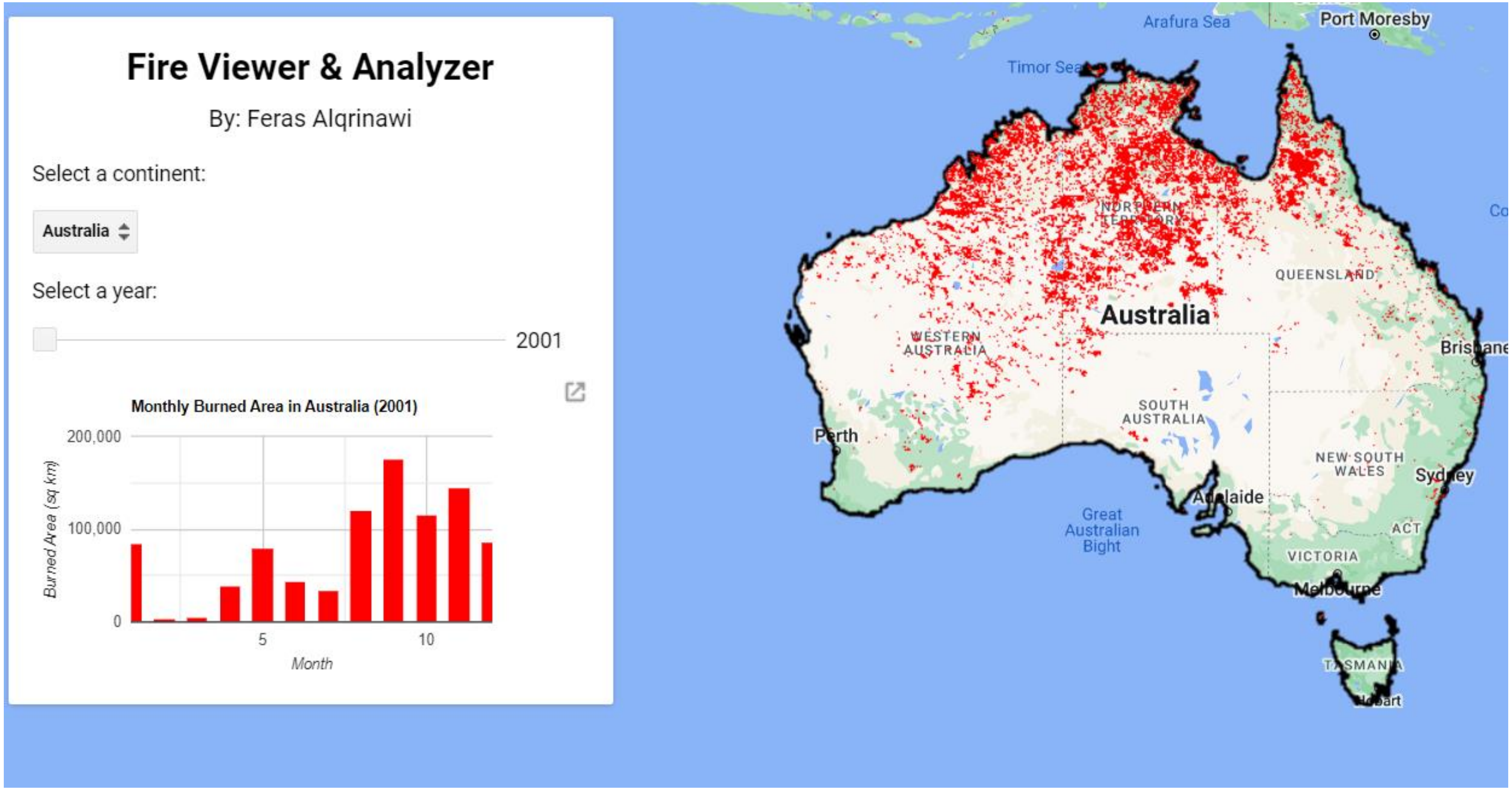
Fire Viewer & Analyzer

Fire Viewer & Analyzer
By Feras Alqrinawi

Fire Viewer & Analyzer

Fire Viewer & Analyzer
By Feras Alqrinawi

- Display the spread of fires on each continent.
- Create interactive graphs showing the fire-affected areas in sq.km for each month throughout the year.
- View and export these graphs as tables (CSV) or images (PNG).



References

•**Karagiannis, G. M., Synolakis, C. E. (2024)**, Wildfire risk management in the era of climate change, *PNAS Nexus*, 3(5), pgae151. doi:10.1093/pnasnexus/pgae151. This study highlights the impact of climate change, urban development in wildland-urban interfaces, and historical fire suppression practices on wildfire risk.

•**Viedma, O., Urbieta, I. R., Moreno, J. M. (2018)**, Wildfires and the role of their drivers are changing over time in a large rural area of west-central Spain, *Scientific Reports*, 8(1), 1-13. Link to article. This research discusses how environmental characteristics like topography and vegetation types influence wildfire probability and severity over time.

•**Viedma, O., Quesada, J., Torres, I., De Santis, A., Moreno, J. M. (2015)**, Fire severity in a large fire in a Pinus pinaster forest is highly predictable from burning conditions, stand structure, and topography, *Ecosystems*, 18(2), 237-250. Link to article. This paper explores how fire severity is influenced by pre-fire vegetation, burning conditions, and topography.

•**Urbieta, I. R., Franquesa, M., Viedma, O., Moreno, J. M. (2019)**, Fire activity and burned forest lands decreased during the last three decades in Spain, *Annals of Forest Science*, 76(3), 1-13. Link to article. This study examines the trends in fire activity and how different vegetation types affect fire spread and severity.

•**A global scale study** of the factors affecting wildfires' distribution over the past 10 years using the second apprach . Link

•**Advances in the study of global forest wildfires** (2021), *Journal of Soils and Sediments*. Link to article. This article reviews the development of forest wildfire research, highlighting the main journals, authors, and research categories involved in this field over the past 30 years.

Thank you!