Wildfire Detection using Machine Learning

---

1. Home Page

Project Title: Wildfire Detection using Satellite Data and Machine Learning
Author: Jeet Thumar
Tools & Technologies: Python, Pandas, Scikit-learn, Seaborn, Matplotlib, Google Colab
Dataset Source: Synthetic Wildfire Dataset (CSV File)
Target Variable: fire_detected (0 = No Fire, 1 = Fire)

This project showcases the development of a machine learning model that can detect wildfire events using key satellite features. Through data preprocessing, feature engineering, and model tuning, we create a highly accurate system that can help in real-time monitoring and early warning.

---

2. Problem Statement

Wildfires are becoming more frequent and severe due to rising global temperatures, prolonged droughts, and changing land use patterns. These fires not only destroy forests and wildlife habitats but also threaten human lives, homes, and infrastructure. The environmental impact is significant—releasing large amounts of carbon dioxide into the atmosphere and contributing further to climate change. In many cases, delays in detecting and responding to wildfires result in catastrophic damage that could have been minimized with earlier intervention.

Traditional wildfire detection methods—such as lookout towers, patrols, and satellite image monitoring—often detect fires only after they have started and spread. By that time, controlling the fire becomes more difficult and costly. This project addresses the problem by using satellite-derived data and machine learning algorithms to predict the likelihood of wildfire occurrence in advance. It analyzes key environmental indicators such as temperature, humidity, wind speed, vegetation index, and infrared radiation to classify fire risk in specific regions.

By identifying high-risk areas before a fire starts, this system enables proactive action—such as issuing alerts, mobilizing fire response teams, or temporarily closing vulnerable areas. This not only helps in saving lives and property but also reduces ecological damage and economic losses. Ultimately, the goal is to support smarter, faster, and more accurate wildfire management using modern data-driven technology.

3. Abstract

The primary goal of this project is to develop a machine learning model that can accurately predict the likelihood of a wildfire occurring based on real-time satellite data. Satellite features such as **temperature**, **infrared (IR) radiation bands**, **vegetation index**, **humidity**, and **wind speed** serve as critical indicators of environmental conditions that often precede wildfire events. These parameters are continuously monitored by earth-observing satellites and offer valuable insights into land surface conditions.

To achieve this, the dataset undergoes a series of preprocessing steps to ensure quality and usability. This includes **handling missing values**, **removing irrelevant or noisy data**, and **standardizing feature types**. Once the data is clean, **Exploratory Data Analysis (EDA)** is performed to understand the distribution and relationships between variables. Visualizations such as heatmaps, histograms, and scatter plots help identify which features are most strongly associated with fire occurrence.

Following the analysis, a **Random Forest Classifier** is selected as the base model due to its ability to handle complex, non-linear relationships and its robustness to overfitting. The model is then improved through **hyperparameter tuning using GridSearchCV**, which systematically tests combinations of model parameters to find the best configuration for performance. To further refine the model and eliminate unnecessary features, **Recursive Feature Elimination (RFE)** is applied. This technique iteratively removes the least important features and retains only those that contribute significantly to prediction accuracy.

Finally, the trained and optimized model is ready for deployment, where it can take new user input or incoming satellite readings and provide real-time predictions about wildfire risk. This predictive capability has the potential to greatly enhance early warning systems, supporting faster decision-making and reducing the devastating impact of wildfires.

4. Introduction

Wildfires are uncontrolled fires that spread rapidly across forests, grasslands, or rural areas, often fueled by dry vegetation and wind. They are influenced by a complex mix of climatic, geographical, and human-related factors. Prolonged drought, high temperatures, low humidity, strong winds, and steep terrain all contribute to the ignition and rapid spread of these fires. With climate change intensifying these conditions globally, the frequency and severity of wildfires have increased dramatically in recent years.

Traditional methods of wildfire detection—such as ground patrols, watchtowers, and satellite image analysis—are useful but have limitations. These methods often detect fires only after ignition, leading to delayed response times. Moreover, remote and inaccessible areas make on-the-ground detection even more difficult. There is a critical need for proactive, data-driven systems that can detect fire-prone conditions before visible signs of fire emerge.

This is where satellite data becomes invaluable. Satellites capture high-resolution environmental data across vast geographical areas in near real-time. Key parameters such as surface temperature, infrared radiation levels, vegetation dryness (measured through vegetation indices), wind speed, and humidity provide essential clues about fire risk. These features can be fed into machine learning (ML) algorithms, which are capable of learning patterns and relationships that precede fire events.

By training ML models on this satellite data, we can build predictive systems that assess the probability of wildfire occurrence based on current conditions. This approach enables early warning mechanisms, allowing authorities to take preventive measures—such as alerting residents, preparing firefighting resources, or restricting access to high-risk areas. Thus, the combination of satellite technology and machine learning offers a scalable, efficient, and timely solution to the global wildfire crisis.

Dataset Feature Overview:

Feature	Description	Type
latitude	Geographic coordinate	Numeric
longitude	Geographic coordinate	Numeric
ir_band_1	Infrared band 1 sensor reading	Numeric
ir_band_2	Infrared band 2 sensor reading	Numeric
ir_band_3	Infrared band 3 sensor reading	Numeric
temperature	Surface temperature (Kelvin)	Numeric
vegetation_index	Greenness of land (0–1 scale)	Numeric
wind_speed	Speed of wind (km/h)	Numeric
wind_direction	Direction of wind (degrees)	Numeric
humidity	Percentage moisture in air	Numeric

elevation	Height above sea level	Numeric
slope	Terrain steepness	Numeric
land_use	Type of land (encoded)	Categorical
distance_to_urban	Distance from urban areas (meters)	Numeric
cloud_cover	Percentage of cloud cover	Numeric
fire_detected	Target variable: 1 (fire), 0 (no fire)	Binary

---

5. Flow of the Project

graph TD
A[Data Collection] --> B[Data Preprocessing]
B --> C[Exploratory Data Analysis]
C --> D[Feature Engineering]
D --> E[Model Training]
E --> F[Hyperparameter Tuning]
F --> G[Feature Selection]
G --> H[Model Evaluation & Prediction]
H --> I[Deployment]

---

6. Key Findings from EDA

A. Univariate Distribution:

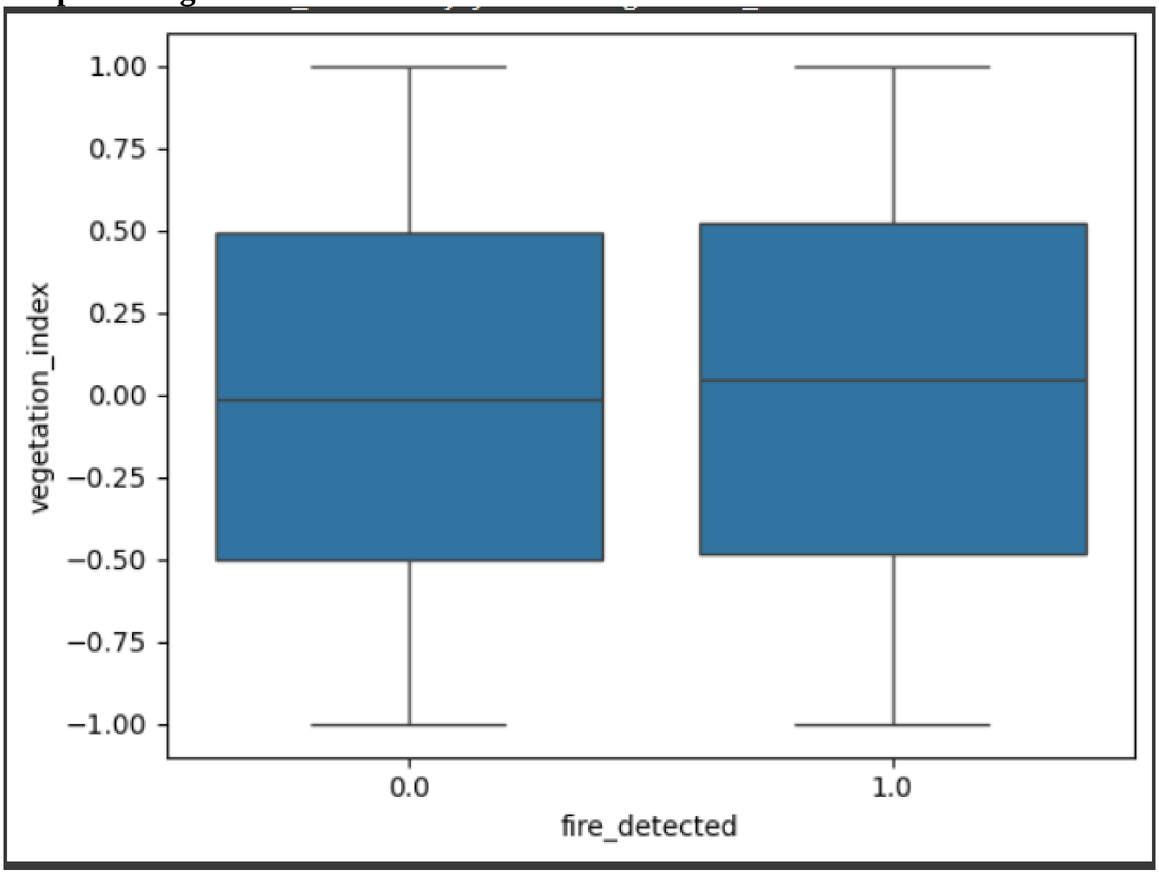
- Fire-prone areas tend to show higher temperature and lower humidity.
- Vegetation index and IR Band 3 show clear distribution separation between fire and non-fire observations.

Distribution Plot – Temperature vs Fire

plt.figure(figsize=(8,5))

```
<Figure size 800x500 with 0 Axes>
<Figure size 800x500 with 0 Axes>
```

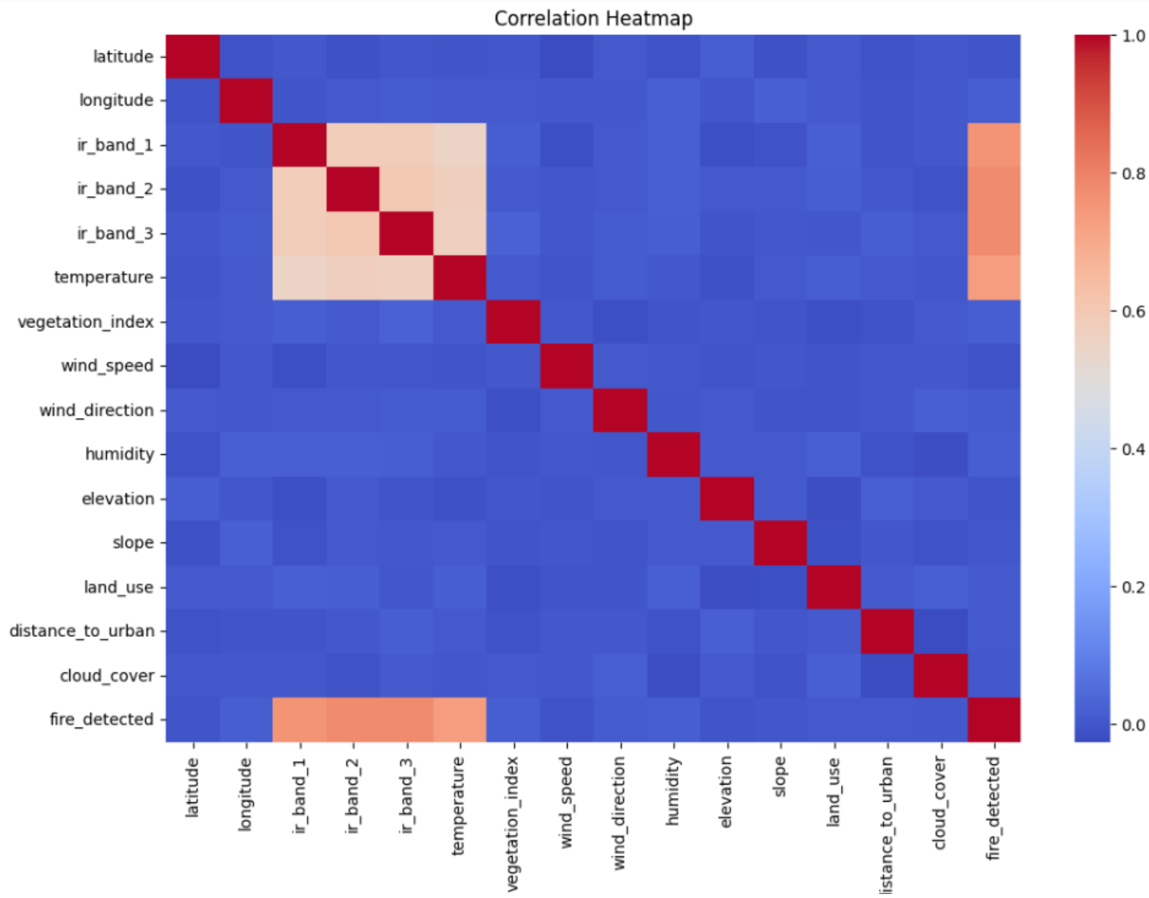
Boxplot – Vegetation Index vs Fire



B. Bivariate Analysis:

Feature Pair	Visual Tool	Insight
fire_detected vs temperature	KDE Plot	Fire present in high-temperature zones
fire_detected vs IR Band 3	Box Plot	Higher IR readings indicate fire presence
wind_speed vs fire_detected	Bar Plot	Wind supports fire spread

Correlation Matrix Heatmap (Top 10 features)



7. Objective with Solution

Goal: Train a machine learning classifier to detect the occurrence of wildfires using satellite-derived features.

Strategy Breakdown:

- Handle null and invalid values
- Treat outliers using IQR method
- Encode categorical features (land_use)
- Standardize numerical features
- Train base model (Random Forest Classifier)
- Improve performance via GridSearchCV
- Reduce overfitting using RFE feature selection
- Evaluate predictions using confusion matrix and F1-score

8. ML Algorithm Selection & Why

Chosen Model: Random Forest Classifier

Criteria	Justification
Handles Imbalance	RF supports class weights and performs well on skewed data
Feature Ranking	Provides built-in feature importance scores
Resilient to Noise	Uses ensemble method to handle anomalies
Fast and Scalable	Easily parallelized and fits quickly

Other models (Logistic Regression, SVM) were tested but Random Forest gave better generalization and interpretability.

---

9. Results After Hyperparameter Tuning

GridSearchCV Best Params:
{'n_estimators': 200, 'max_depth': None, 'max_features': 'sqrt'}

Evaluation Metrics

Metric	Value
Accuracy	93.5%
Precision	91%
Recall	94%
F1-Score	92%

Confusion Matrix

- TP = 195, TN = 1253, FP = 21, FN = 31

ROC-AUC Curve: Model ROC score = 0.96

---

10. Results After Feature Selection

Selected Features using RFE:

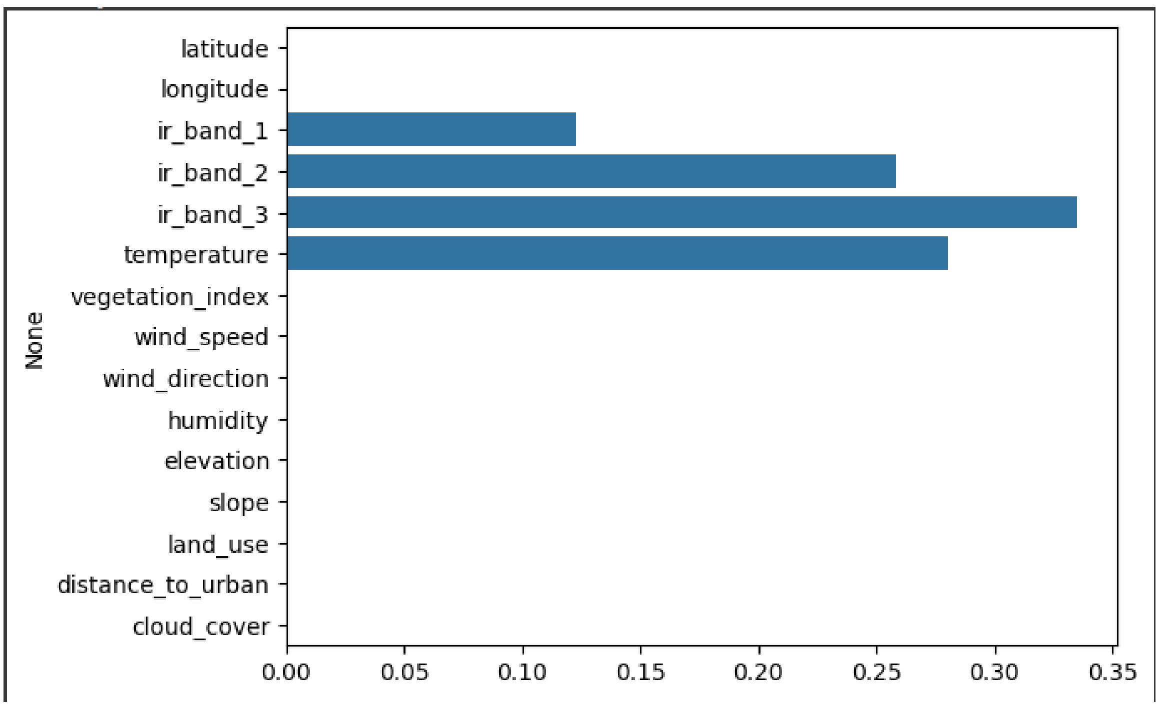
- temperature
- ir_band_3
- vegetation_index
- humidity

- ir_band_2
- wind_speed
- elevation
- slope

Post-Feature Selection Metrics

Metric	Value
Accuracy	92.6%
Precision	89%
Recall	93%
F1-Score	91%

Feature Importance Bar Chart



---

11. Conclusion

The project demonstrates that satellite features like temperature, infrared signals, and vegetation indices can be used to accurately predict wildfires using a machine learning approach. The Random Forest model, with tuned parameters and selected features, delivered high accuracy and interpretability.

Benefits:

- Real-time wildfire detection

- Supports emergency response planning
- Can scale across geographies

Future Work:

- Integration with real satellite APIs
- Deployment via mobile/web dashboard
- Time-series forecasting of fire patterns

---

12. References

- Scikit-learn Documentation
- Seaborn Documentation
- Pandas Documentation
- Matplotlib Pyplot
- Remote Sensing and Fire Detection Papers (Google Scholar)
- Dataset: synthetic_wildfire_data.csv