# **Wildfire Footprints: Cloud-Based Forecasting with NASA MODIS Data**

**Team Members**:  
 Ritu Patel & Dhwanil Mori

GITHUB LINK:

<https://github.com/Ritu0927/Cloud_Computing>

## **1. Introduction**

This project focuses on forecasting wildfire activity across the United States using **NASA MODIS satellite data** from **2003, 2013, and 2023**.  
 We leveraged **AWS Cloud Services** including **EC2**, **RDS**, and **Jupyter Notebook** to build an end-to-end cloud-native data pipeline and developed predictive models to analyze both **Fire Radiative Power (FRP)** and **Brightness** trends.

A screenshot of a computer

AI-generated content may be incorrect.

## **2. Project Scope & Motivation**

* Analyze wildfire trends over two decades (2003–2023).
* Forecast wildfire intensity (FRP) and heat magnitude (Brightness) using statistical and deep learning models.
* Motivation: Provide data-driven insights for wildfire risk management and climate resilience.

## **3. Cloud Architecture Overview**

| Step | Task | Cloud / Local | Notes |
| --- | --- | --- | --- |
| 1 | Upload CSVs to EC2 | AWS (Cloud) | Jupyter Notebook Server |
| 2 | Create RDS MySQL DB | AWS (Cloud) | Wildfire data storage |
| 3 | Load Data into RDS | AWS (via Jupyter) | Python ingestion scripts |
| 4 | Analyze and Model | AWS (Jupyter) | VAR and LSTM models |
| 5 | Entirely Cloud-Native | AWS | No local machine required |

## 

## **4. Data Sources & Key Features**

* **Primary Source**: NASA MODIS (Terra and Aqua satellites)
* **Key Variables**:  
  + Fire Radiative Power (FRP)
  + Brightness
  + Confidence score
  + Latitude and Longitude
* **Dataset Size**: Approximately 100,000 records per year
* **Years Considered**: 2003, 2013, 2023

## **5. Feature Engineering**

* Temporal features: Month, Year extraction
* Categorized Fire Intensity (low, medium, high based on FRP)
* Normalization of Brightness values
* Input sequence preparation for time-series forecasting

A graph of a box plot

AI-generated content may be incorrect. A graph with a line

AI-generated content may be incorrect. A purple rectangular object with white text

AI-generated content may be incorrect. A graph with a blue rectangle

AI-generated content may be incorrect.

## **6. Data Processing Pipeline**

* Upload raw CSVs to AWS EC2
* Store structured data in AWS RDS MySQL
* Connect Jupyter Notebook to RDS for Exploratory Data Analysis (EDA)
* Conduct feature engineering
* Prepare datasets for model training

*A screenshot of a computer

AI-generated content may be incorrect.*

*A screenshot of a computer

AI-generated content may be incorrect. A screenshot of a computer

AI-generated content may be incorrect.*

## **7. Forecasting Approaches**

**1. VAR (Vector AutoRegression)**:

* Classical statistical model assuming linear relationships.
* Captures dependencies between multiple variables over time.

**2. LSTM (Long Short-Term Memory Network)**:

* Deep learning model specialized for sequential data.
* Captures complex non-linear patterns and long-term dependencies.

## **8. Model Results & Metrics**

**Performance Comparison for Brightness and FRP Forecasting**:

A black and white table with white text

AI-generated content may be incorrect.**Key Observations**:

* **VAR** produced smoother, more reliable forecasts for Brightness.
* **LSTM** better captured FRP volatility but suffered from overfitting.

## **9. Comparative Forecast Analysis**

A graph with lines and numbers

AI-generated content may be incorrect.

A graph of different colored lines

AI-generated content may be incorrect.

* **VAR**:  
  + Works well for smoother signals (Brightness).
  + Generates consistent, low-volatility forecasts.
* **LSTM**:  
  + Better at capturing complex seasonal patterns.
  + Tends to overfit extreme FRP peaks.

Overall, VAR produced more interpretable outputs for operational use, while LSTM offered deeper pattern capture for research exploration.

## 

## **10. Challenges Faced**

* **Data Quality**:  
   Handling missing values and noisy wildfire detections.
* **Cloud Setup**:  
   Securely configuring EC2, RDS, and Jupyter connections.
* **Forecasting Volatility**:  
   Difficulty in predicting FRP due to its unstable, spiky nature.

## **11. Future Work**

* **Migrate workflows** to Amazon SageMaker for scalable model training and tuning.
* **Automate real-time data ingestion** using AWS Lambda and Amazon Kinesis.
* **Archive historical data** using Amazon S3 Glacier for cost optimization.
* **Deploy models as APIs** using AWS Elastic Beanstalk or API Gateway.
* **Enhance security and monitoring** with AWS IAM, Secrets Manager, and CloudWatch.