MIT | Academy of Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

WildFire Prediction

TY B.Tech
Computational Intelligence Project Report

SUBMITTED BY

Pratik Bhosale – 202301070147

Sameer Kadam – 202301070149

Kartikya Gawali – 202301070151

Yash More – 202301070153

GUIDED BY
Dr. Abhilasha Joshi
Prof. Department of E&TC

MIT ACADEMY OF ENGINEERING, ALANDI (D), PUNE-412105

MAHARASHTRA (INDIA)

Nov, 2026

ACKNOWLEDGEMENT

We express our sincere gratitude to all those who supported and contributed to the successful completion of this project. We extend our deep appreciation to our respected guide, **Dr. Abhilasha Joshi**, for her continuous guidance, expert insights, and valuable encouragement throughout the duration of this work. Her mentorship played a crucial role in shaping the direction and quality of this project.

We also thank the Dean and the Department of Electronics & Telecommunication Engineering for providing the necessary facilities, technical resources, and academic environment that enabled the smooth execution of this project.

We are grateful to our peers and colleagues for their cooperation, constructive suggestions, and constant motivation at every stage of this work. Lastly, we acknowledge all individuals and institutions whose direct or indirect contributions helped us complete this project successfully.

ABSTRACT

Wildfires have become increasingly severe in recent years, particularly in California, causing substantial environmental destruction and economic loss. Predicting wildfire severity at an early stage can significantly aid emergency response teams in effective resource management and timely intervention.

This project presents a machine learning–based approach to predicting wildfire severity, measured in terms of **AcresBurned**, using various incident-related features such as location, containment percentage, and firefighting resources deployed. The dataset undergoes thorough preprocessing steps including missing value imputation, feature encoding, normalization, and outlier handling. Exploratory data analysis (EDA) identifies key relationships between variables such as containment level, resource allocation, and final burn area.

Multiple regression models including **Linear Regression, Decision Tree, Random Forest, and XGBoost** are trained and evaluated. Among these, the **Random Forest** model demonstrates the best predictive performance with high accuracy and robustness.

The final system provides a data-driven and reliable method to estimate wildfire severity, supporting early risk assessment and resource planning. This approach can be integrated into real-time monitoring dashboards for wildfire management authorities.

1.Dataset Overview

The dataset used in this project contains more than 1,600 wildfire incident records from California, with features describing location, containment status, and firefighting resources.

Key Features:

- **County** (categorical)
- **Latitude**, **Longitude**
- **PercentContained**
- **PersonnelInvolved**
- **Engines**, **Helicopters**
- **Fire Discovery Time**, **Control Time** (optional)

Target Variable:

- **AcresBurned** — Continuous numerical variable representing wildfire severity.

These features collectively help predict how large a wildfire may grow based on initial conditions.

2 .Data Preprocessing Steps:

a. Handling Missing Values

Wildfire incident datasets often contain missing values due to incomplete reporting or delayed data collection.

Approach Used:

- Numerical fields → **Median Imputation**
- Categorical fields → **Most Frequent Imputation**

This preserves data distribution while minimizing distortion.

b. Encoding Categorical Variables

- **County** → Label Encoding / One-Hot Encoding

c. Outlier Treatment

Outliers frequently occur in variables such as **AcresBurned**, **PersonnelInvolved**, and **Engines**.

Method:

- IQR Method for resource-related features
- Capping extreme values for burn area

d. Feature Scaling

Scaling is important for models like Linear Regression.

Scaler Used: **StandardScaler()**

- Mean = 0, Standard Deviation = 1

e. Train–Test Split

- **80% Training**, **20% Testing**
- Randomized split to ensure fair evaluation

3. Exploratory Data Analysis (EDA)

Key Observations:

a. Containment vs Severity

- Lower **PercentContained** at early stage correlates with extremely high burn area.

b. Resource Allocation Trends

- Wildfires with more **personnel and engines** deployed early tend to have lower final burn area.

c. Geographic Patterns

- Certain counties consistently show higher burn severity due to terrain and climate.

d. Feature Correlations

Strong correlations were found between:

- **PersonnelInvolved ↔ AcresBurned**
- **Engines ↔ AcresBurned**
- **PercentContained ↔ AcresBurned** (negative correlation)

4. Model Development Summary

The following regression models were trained and evaluated:

- **Linear Regression**
- **Decision Tree Regressor**
- **Random Forest Regressor**
- **XGBoost Regressor**

Evaluation Metrics:

- **MAE (Mean Absolute Error)**
- **RMSE (Root Mean Squared Error)**
- **R^2 Score**

These metrics help determine accuracy, consistency, and overall predictive capability.

5. Best Model: Random Forest

Reasons for Selection:

- Handles complex, non-linear patterns effectively
- Performs well on mixed data types
- Resistant to overfitting due to ensemble averaging
- Provides meaningful feature importance values
- Achieves lowest MAE and RMSE among all models

Random Forest offers the most balanced performance and reliability for predicting wildfire severity.

6. Results & Interpretation

- containment drastically reduces severity.
- **PersonnelInvolved** and **Engine count** play significant roles in limiting spread.
- Random Forest's predictions align closely with actual burn area values.
- High R^2 score indicates strong predictive accuracy.

The model **PercentContained** shows the strongest influence on wildfire size — early effectively captures real wildfire behavior, making it suitable for decision-making applications.

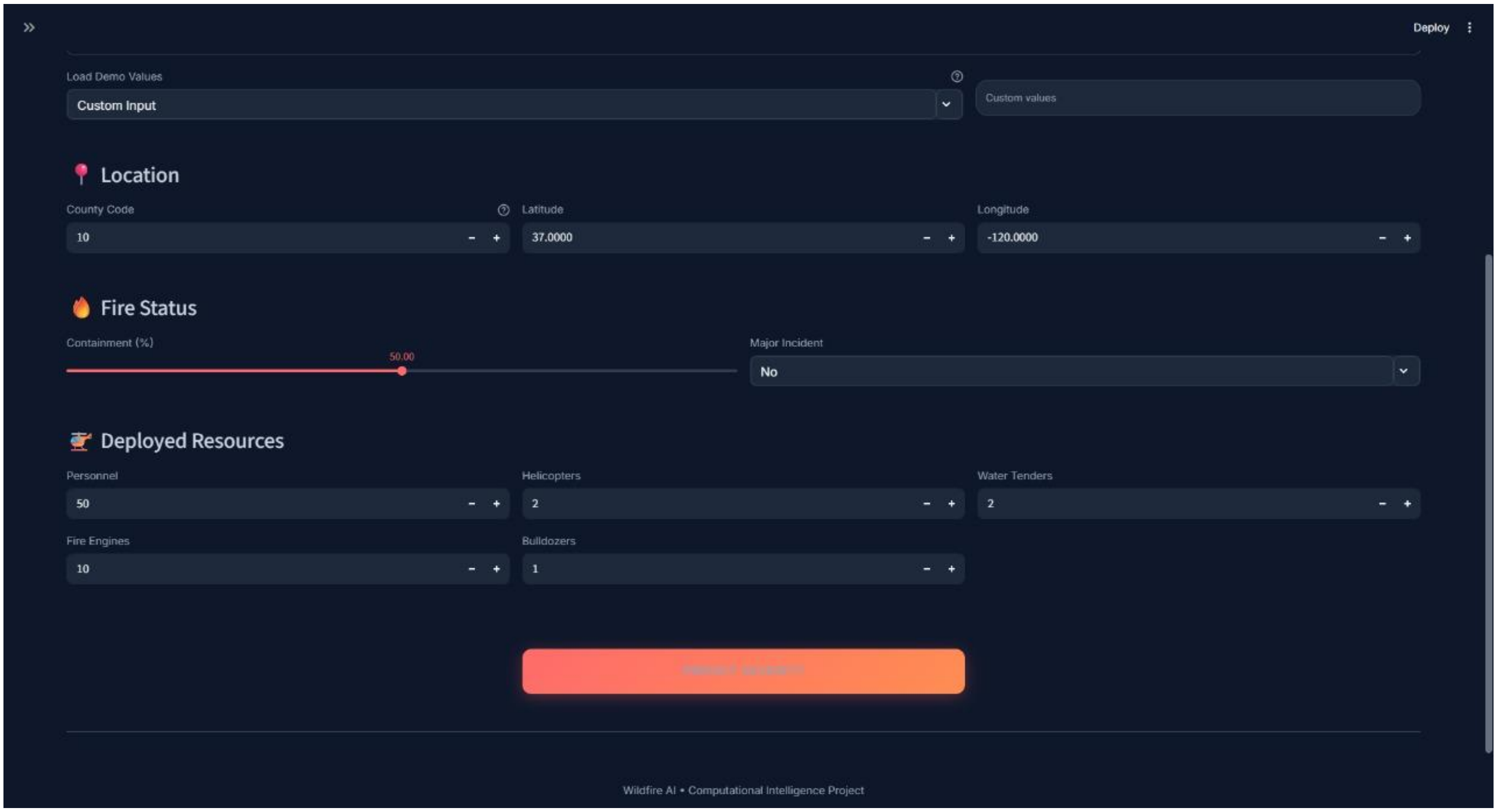
7. Conclusion

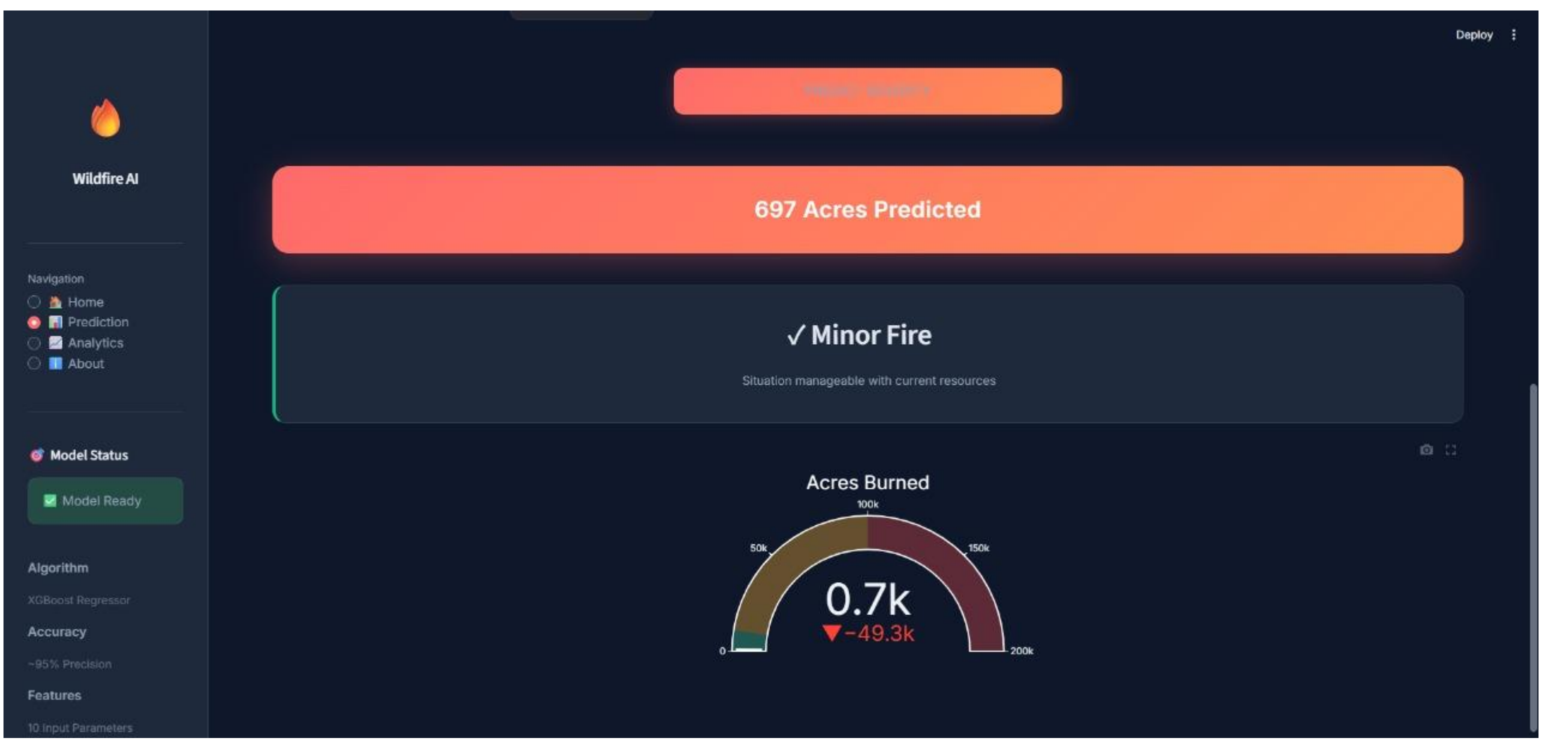
This project demonstrates that machine learning can effectively predict wildfire severity using initial incident features. A combination of resource data, containment status, and location-based attributes provides a strong foundation for modeling wildfire behavior.

Among the models evaluated, **Random Forest** delivers the most accurate and stable predictions. This system can support wildfire authorities in early risk assessment, enabling smarter and faster decision-making.

8. Future Enhancements

- Integrate real-time wildfire data from APIs (NASA FIRMS, Cal Fire)
- Use deep learning models (LSTM/GRU) for time-series predictions
- Deploy scalable cloud-based architecture using FastAPI or AWS
- Build interactive dashboard for real-time wildfire severity monitoring
- Add geospatial analysis and heatmap visualization for risk mapping


Deploy
Load Demo Values
Custom Input
Custom values
Location
County Code
10
Latitude
37.0000
Longitude
-120.0000
Fire Status
Containment (%)
50.00
Major Incident
No
Deployed Resources
Personnel
50
Helicopters
2
Water Tenders
2
Fire Engines
10
Bulldozers
1
Wildfire AI • Computational Intelligence Project


Deploy
Wildfire AI
Navigation
Home
Prediction
Analytics
About
Model Status
Model Ready
Algorithm
XGBoost Regressor
Accuracy
~95% Precision
Features
10 Input Parameters
697 Acres Predicted
✓ Minor Fire
Situation manageable with current resources
Acres Burned
0.7k
▼−49.3k
0
50k
100k
150k
200k