Wildfire Ignition Risk Prediction(Colorado)

Malik Ibrahim Ali Khan

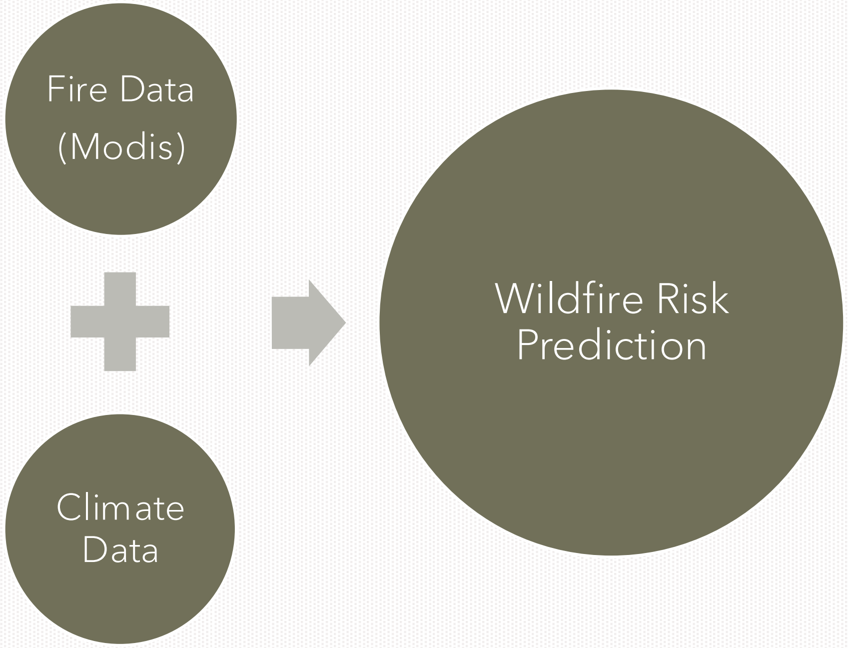
Regis University

Mibrahimalikhan001@regis.edu

Problem Statement and Data Source

- **Problem**: Can Climate data be used to predict wildfire risk in Colorado?

Data set	Description	Source
MODIS FIRMS	Satellite wildfire occurrences.	NASA FIRMS
gridMet	Temperature and Precipitation (daily).	Merced – Climatology Lab(California)



Location	Time	Scale	Granularity
Colorado	2018-2023	35M+	Daily Grid

EDA: Variables, Insights & Modeling Decisions

Variables

Climate

- Precipitation(Daily, Lagged, Rolling)
- Temperature(Max, Min)

Fire

- Binary occurrence (0/1)

Key Insights

- Fire is extremely rare (0.1%)
- Strong Seasonality (Summer)
- Dryness(Rolling precip) is key
- Temperature is secondary

Modeling Decisions

- Use Risked Ranking with lagged & Seasonal features

Data Preparation and Feature Engineering

Data Integration

- Filtered Fire Data to Colorado
- Aligned Fire Data to gridMET
- Created Binary Fire labels
- Combined 2018-2023 dataset

Feature Engineering

- Lagged Precipitation(1, 3, 7 days)
- Rolling Precipitation(7,14 days)
- Day of year (Seasonality)

Modeling Workflow

Raw Data
Modis Firms
gridMET Climate Data

→

Features
Lagged precipitation
Rolling Precipitation
Seasonal features

→

Models
Logistic Regression
Random Forest
Gradient Boosting
XGBoost

↓

Evaluation
Train : 2018-2022
Test- 2023(Unseen Data)

←

Output
Risk Ranking of Locations

Models & Key Hyper Parameters

Logistic Regression

- Class Imbalance Handling

Random Forest

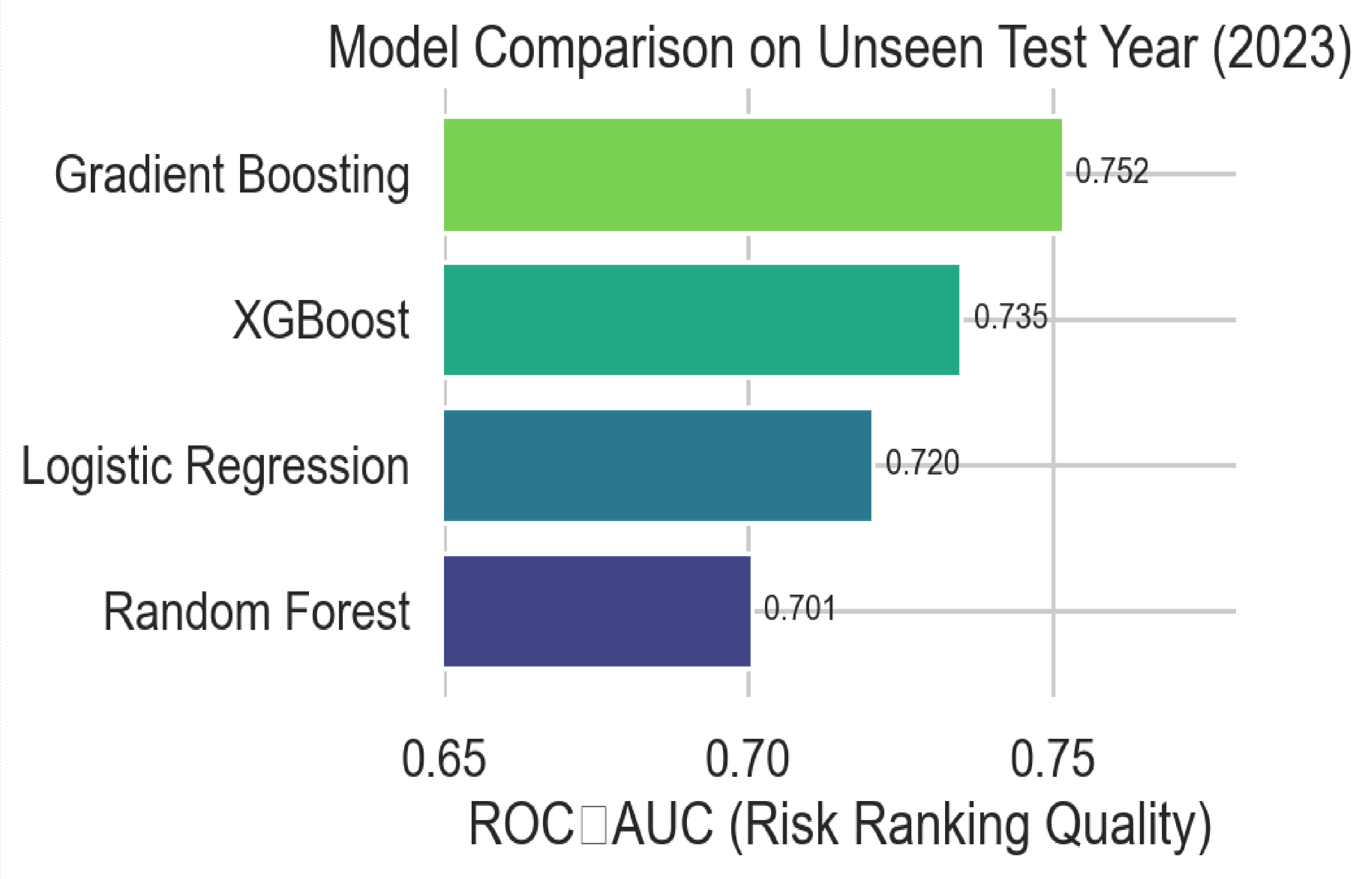
- Tree depth ,Leaf size

Gradient Boosting / XGBoost

- Learning rate and number of trees
- (Subsampling) XGBoost

Key Insight :
Tree-Based Models performed better due to non-linear relationships and feature interactions.

Model Comparison on Unseen Test Year (2023)

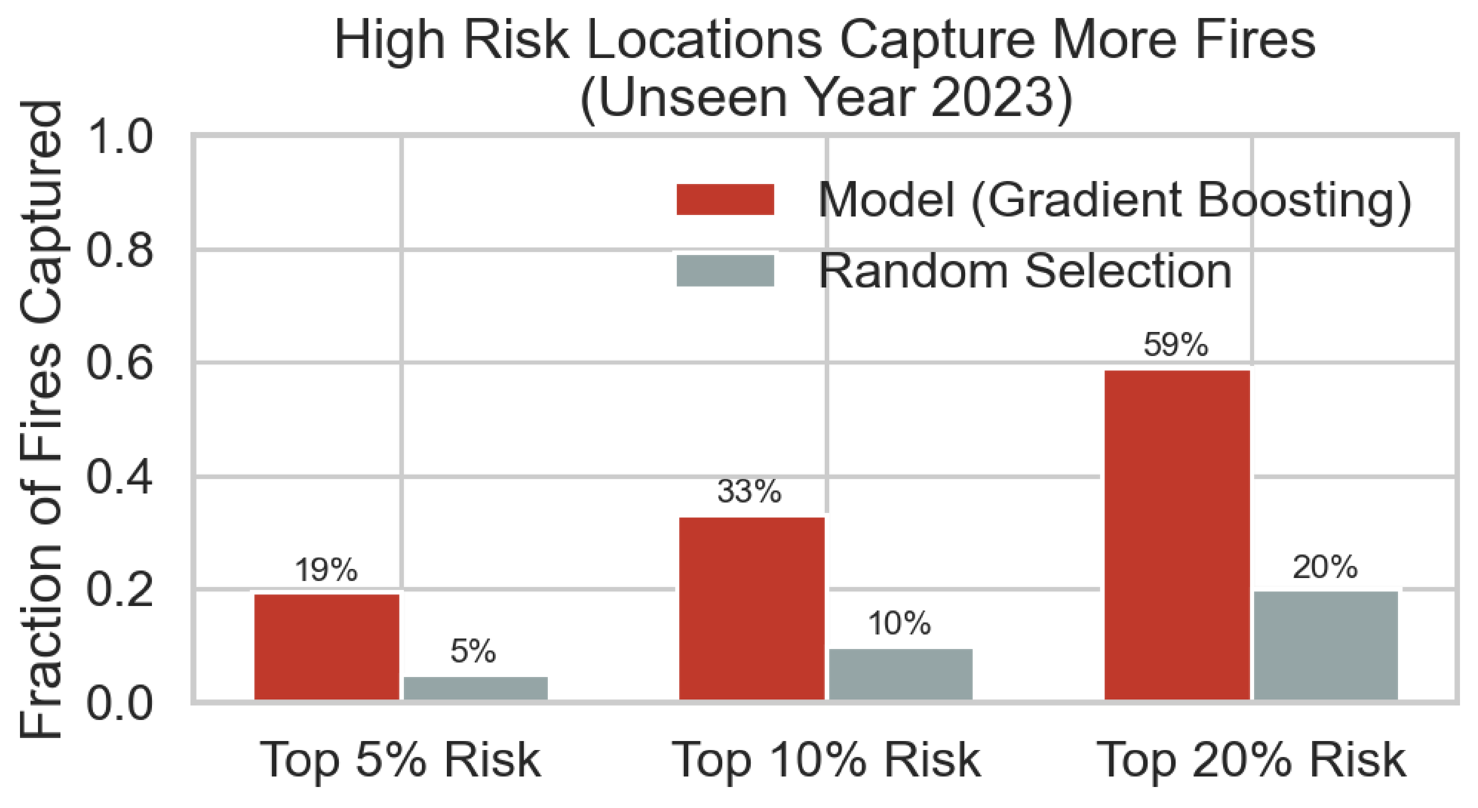


Tree based models outperform Logistic Regression.

Gradient Boosting shows strongest risk-ranking performance

Evaluation performed on unseen year 2023 data

Wildfire Capture on Unseen Year (2023)



Top 5% of locations capture nearly 4x more fire than random

Top 20% of locations capture 60% of Total fires

Performance gap increases as focus Shifts to Higher- Risks areas

Fire Ignition are highly concentrated in a small fraction of locations

Results and Findings

- The Model Identifies a small group of locations where **wildfire ignitions occur far more frequently**
- Climate - based features can **Effectively rank Wildfire Ignition Risk**
- Models Trained on Historical Data **generalize to Unseen future conditions**
- Risk-Ranking provides **practical value or prioritization ,** even when exact prediction is difficult
- Climate variables alone provide **meaningful signal for distinguish higher and lower risk locations.**

Thank you