



JEM CONSULTING

PREDICTING CA WILDFIRES

Wildfire Risk Prediction
& Classification Model



The Problem.

Wildfires in California have been **increasing in frequency and severity** due to climate change, dry conditions, and human activity.

Just from the recent LA wildfires:

29+
Human Lives Lost

16,240
Structures Destroyed

> 1 Megaton
of Carbon Emissions



\$76B+
in property &
capital loss

*"What we're seeing in California right now is more destructive, larger fires burning **at rates that we have historically never seen.**"*

Jonathan Cox, Spokesperson for CAL FIRE



Benefits of Wildfire Prediction.

Accurate wildfire prediction helps **allocate emergency resources** and **inform policymakers**. By identifying high-risk areas in advance, our model enables proactive response, **reducing damages**, **protecting ecosystems**, and **saving lives**.



Faster Emergency Response

Allocate firefighters & resources to high-risk areas.



Better Infrastructure Protection

Prevent damages to homes, roads, and utilities.



Environmental Conservation

Protect forests, wildlife, and air quality.



Public Safety & Awareness

Help communities prepare for evacuation & reduce health risk.

Building the Model.

Our team decided that an **eXtreme Gradient Boosting (XGB) model** is the best suited model for wildfire prediction.

Model building consisted of **3 major phases:** Data Processing, Training & Evaluation, and Optimization.

01. Data Processing

14K+ wildfire records analyzed

Started with 14 features, performed feature selection, handled missing values, and scaled data using `sklearn.preprocessing`.

01.

Model Training & Evaluation

Tested Multiple Models

Evaluated XGB vs. Random Forest, prioritized high recall, and used a train-test split of 80%-20%.

02.

Model Optimization & Visualization

Fine-Tuned for Performance

Used hyperparameter tuning, feature engineering, and SHAP values to refine feature importance.

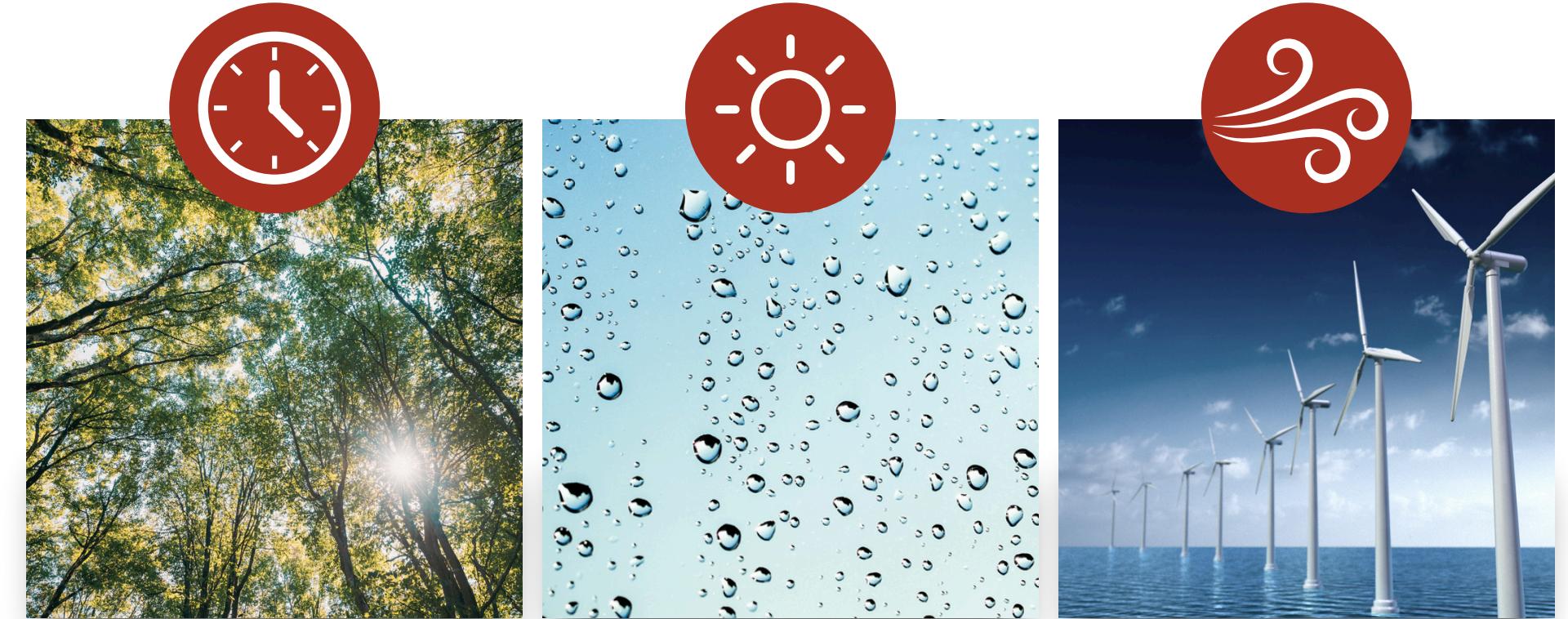
03.

Predictors of Wildfires.

Our model uses **weather-based and time-based features** to **predict** wildfire occurrences.



Day, Month, Season



Wind Speed, Precipitation



Temperature (Min, Max, Range)

By analyzing historical wildfire data, we identified the most influential factors driving fire risk. This allows us to detect patterns and improve model accuracy.

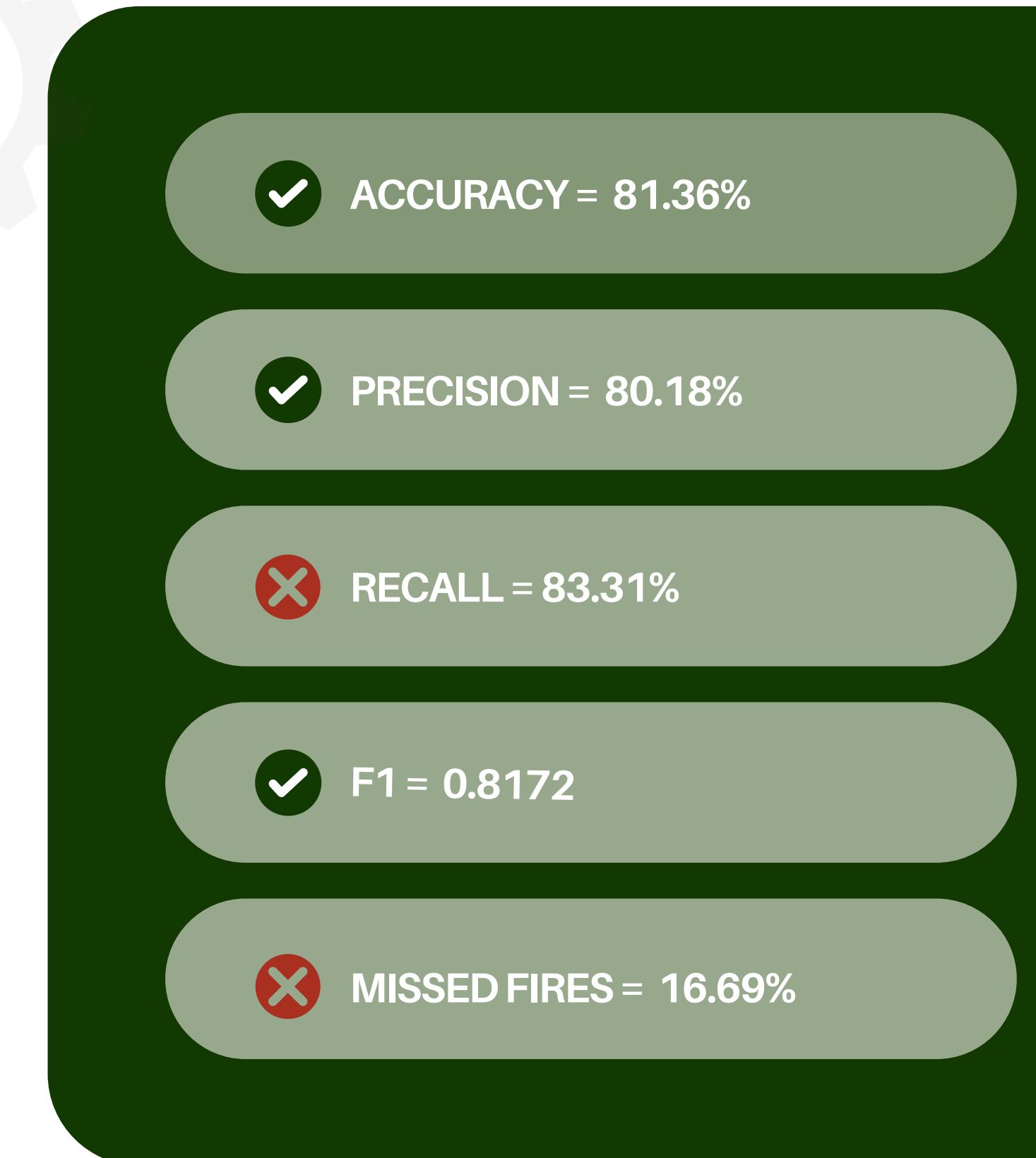
Initial Model Selection.

We evaluated multiple models and **chose XGBoost** for its:

- **High Performance** → Best recall and cross-validation score
- **Handles Imbalanced Data** → scale_pos_weight optimizations
- **Feature Interpretability** → SHAP values highlight key wildfire predictors

Alternative Model Tested: Random Forest

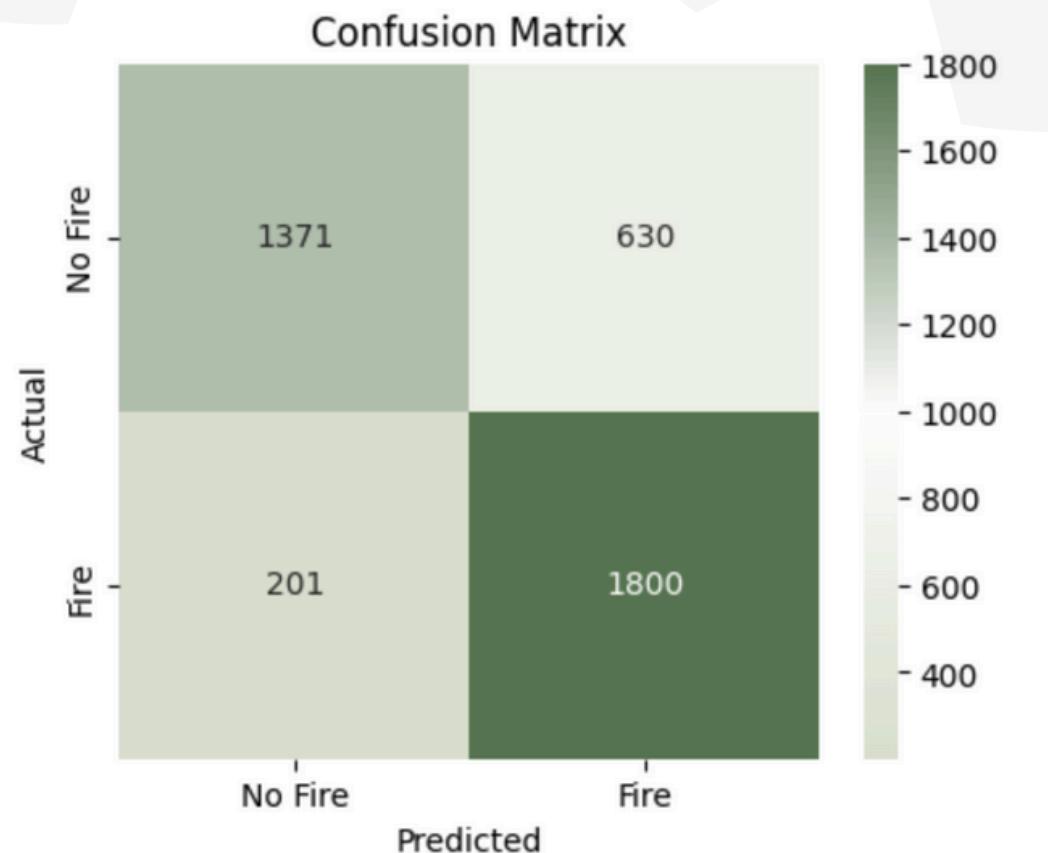
- RF had lower recall (missed more wildfires) but fewer false alarms and a lower cross-validation score.
- XGBoost, after fine-tuning, provided the best balance of recall & precision



Data Source: California Wildfire & Weather Dataset (1984-2025)

Initial XGB Model Evaluation

Training & Evaluation.



Optimizations helped to improve recall.

Reducing the number of missed fires from 325 to 201.

	precision	recall	f1-score
False	0.87	0.69	0.77
True	0.74	0.90	0.81
accuracy			0.79
macro avg	0.81	0.79	0.79
weighted avg	0.81	0.79	0.79

Reducing missed fires came with the tradeoff of precision, leading to an increase in false alarms.

✓ ACCURACY = 79.24%

✓ PRECISION = 74.07%

✓ RECALL = 89.96%

✓ F1 = 0.8125

✓ MISSED FIRES = 10.04%

Optimized XGB Model Evaluation

Optimizing the model.

Hyperparameter Tuning

- Feature Selection
- Introducing Interaction
- RandomCV Search

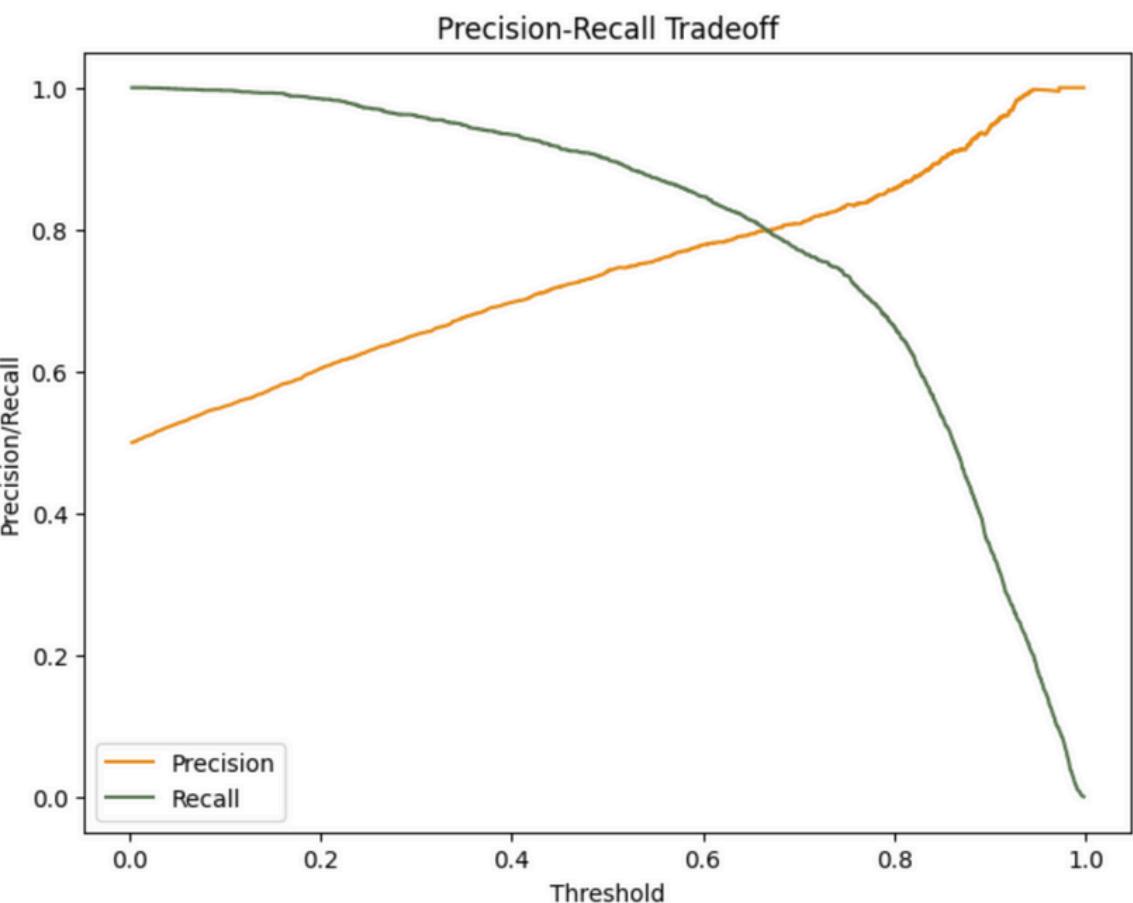
**7.98% increase
in recall**

Decrease Missed Wildfires

Predicting wildfires when they occur is critical, so a focus on decreasing missed wildfires is critical.

**39.82% decrease
in # of missed wildfires**

Precision-Recall Tradeoff



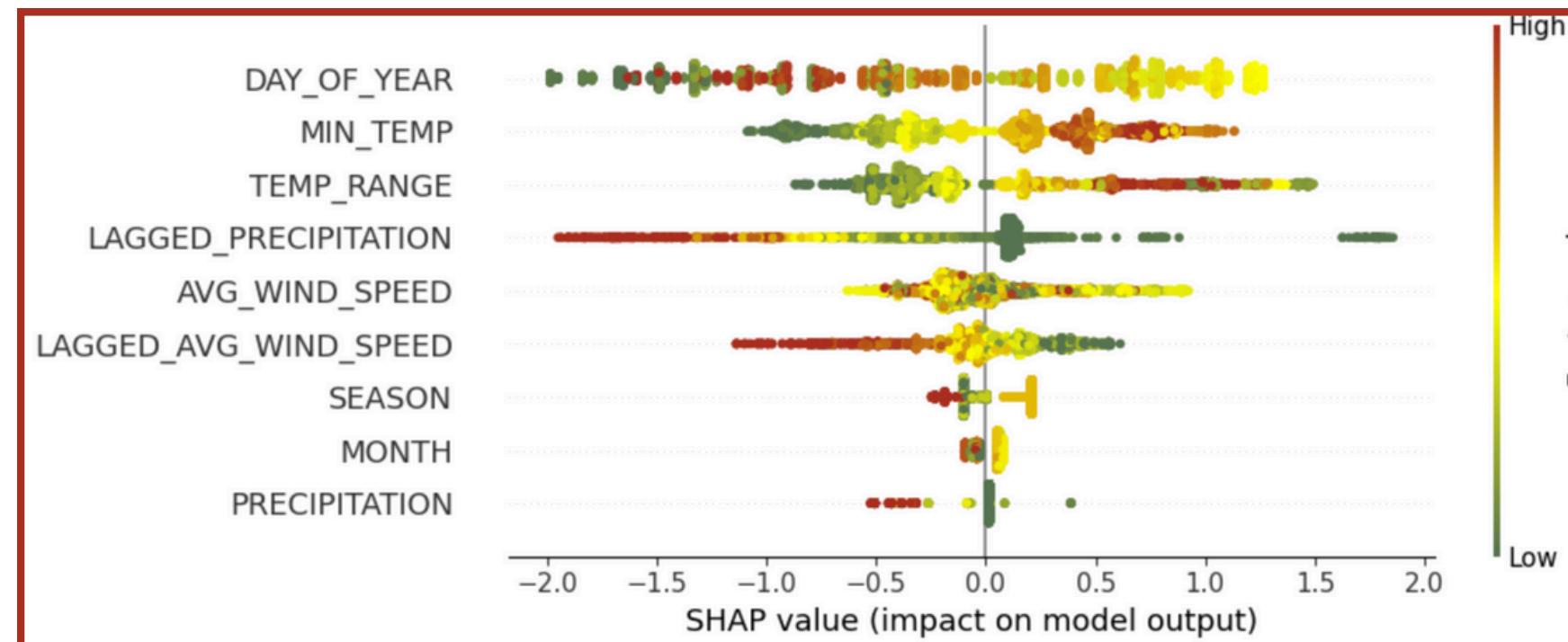
Goal: Decrease false-negatives while ensuring good model performance

Model Decision-Making.

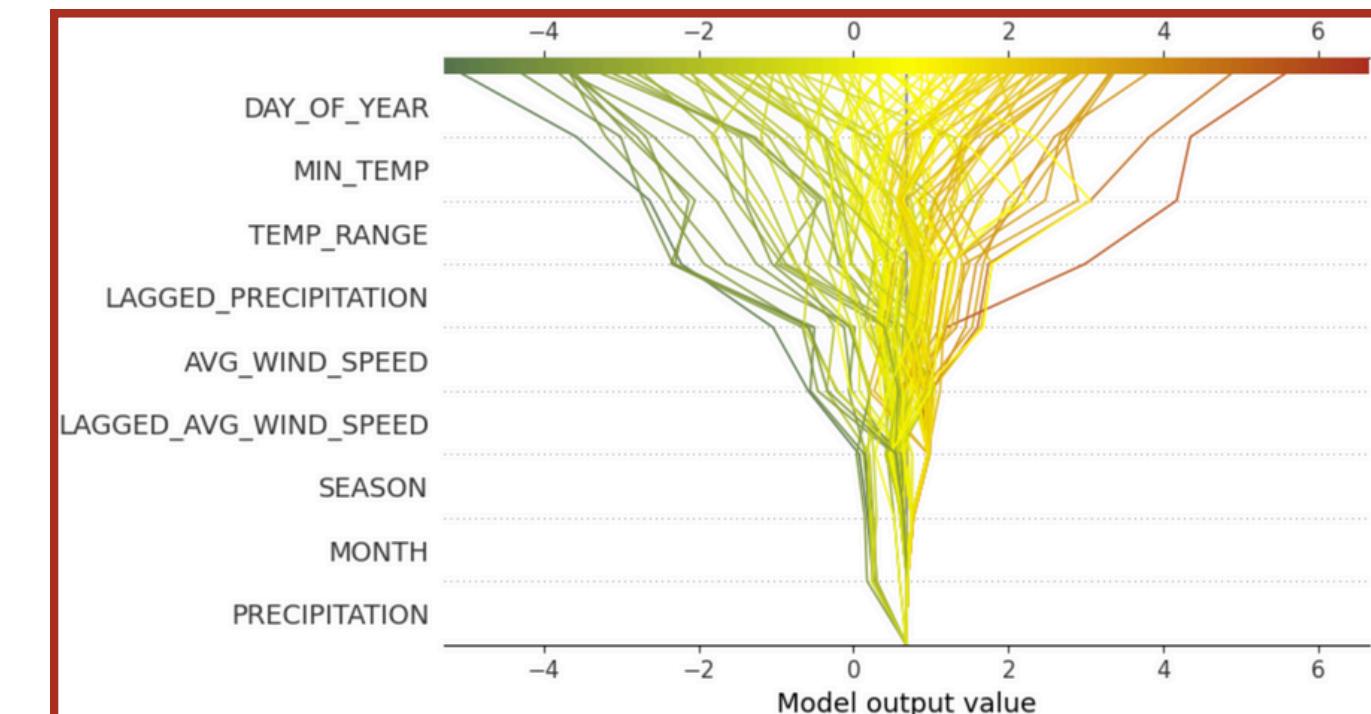
SHAP values explain how much each feature impacts wildfire prediction.

The farther from zero, the stronger the effect.

SHAP Feature Importance: Key Wildfire Drivers



SHAP Decision Path: How Features Influence Predictions



- DAY_OF_YEAR** → Certain times of year have higher risk.
- MIN_TEMP & TEMP_RANGE** → Warmer, fluctuating temperatures increase risk.
- LAGGED_PRECIPITATION** → More rain reduces wildfire probability.
- AVG_WIND_SPEED** → High winds accelerate wildfire spread.



Impact



Fire Departments: Prioritize high-risk areas for faster crew response times.



Homeowners & Insurers: Adjust policies based on predictive fire risk.



Environmental Conservation: Protect forests and reduce carbon emissions.



Government & City Planning: Enforce zoning laws and prevent wildfires in high-risk areas.



JEM CONSULTING

ANY QUESTIONS?

"Only YOU can prevent
wildfires!" - Smokey Bear

