UTM
UNIVERSITI TEKNOLOGI MALAYSIA

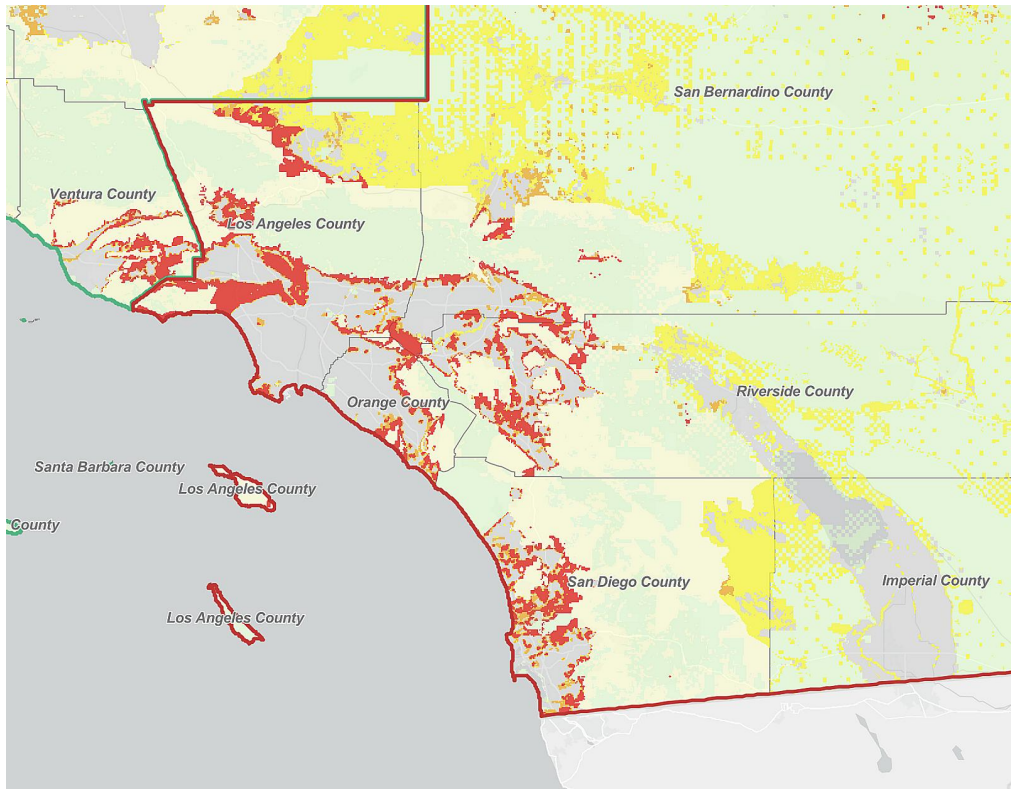
ANALYSIS OF CALIFORNIA'S DECADAL WILDFIRES: THE CONSTRUCTION OF RISK MANAGEMENT MODELS

Presented by: Yi Xindie

Student ID: MCS241027
Supervisor: Prof.Dr.Mohd Shahizan bin Othman
Faculty of computing
Date: 28.06.2025

Innovating Solutions

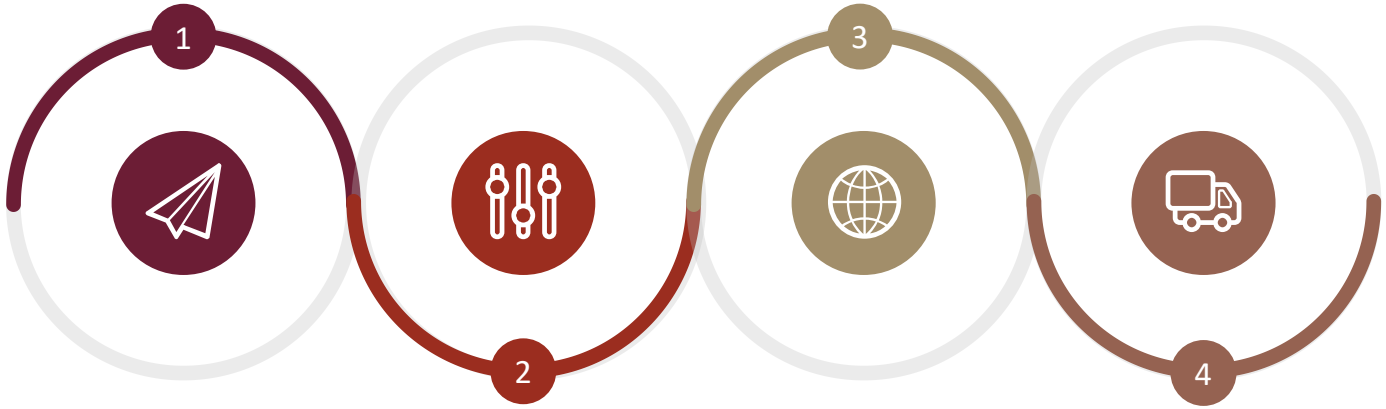
INTRODUCTION



- Wildfires are becoming more frequent and severe in many parts of the world.
- California is one of the regions most affected by wildfires every year.
- Factors like climate change, human activities, and local weather make fire prediction more challenging.
- Many traditional methods struggle to capture these complex risk factors.
- This study aims to explore how machine learning can help predict wildfire risk more effectively.

RESEARCH OBJECTIVES

1. To identify key factors that influence wildfire risk in California.
2. To compare different machine learning models for predicting fire occurrence
3. To develop a practical prediction model based on real wildfire data.
4. To provide suggestions that may help improve wildfire risk management.



METHODOLOGY OVERVIEW

※This project uses a structured workflow:

- ☑ Data Collection
- ☑ Data Preprocessing
- ☑ Model Selection and Training
- ☑ Performance Evaluation
- ☑ Result Interpretation

※Four models were tested: Logistic Regression, SVM, Random Forest, and XGBoost(√).

※Final model performance is compared using accuracy, precision, recall, F1-score, and AUC.

DATA & VARIABLES

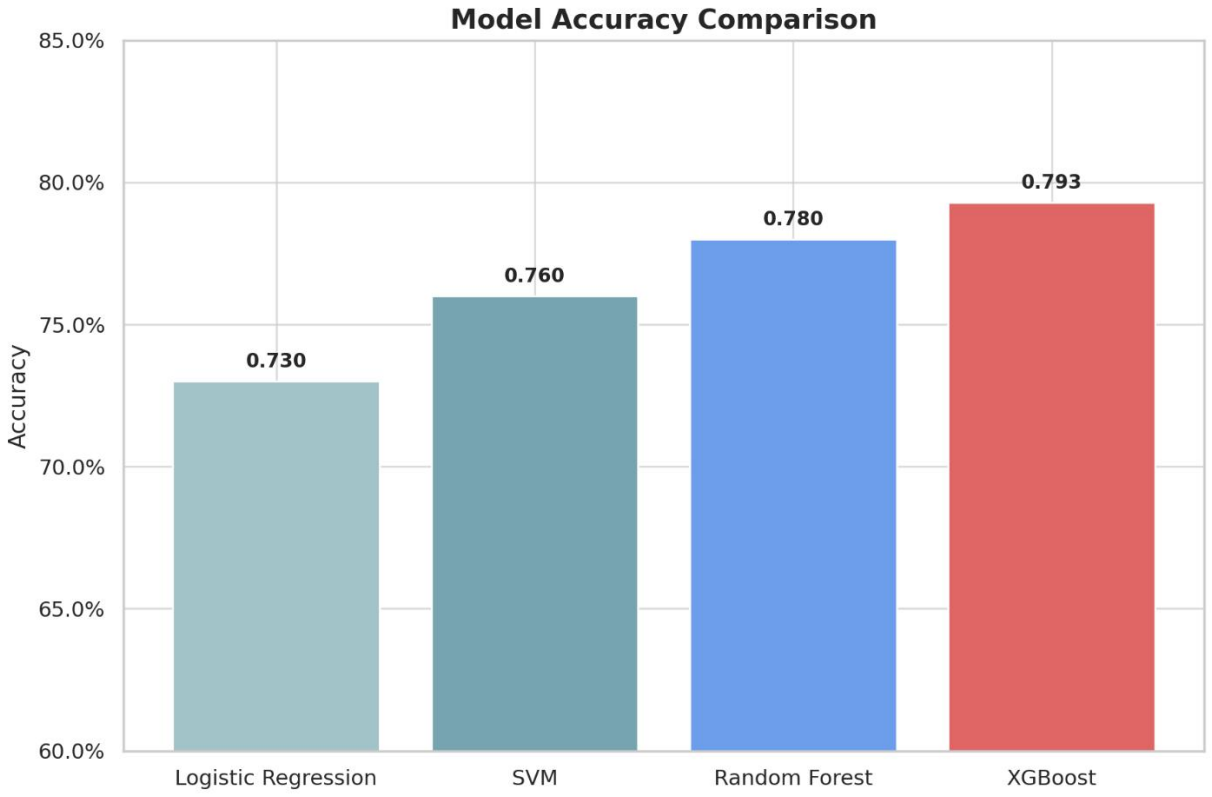
The dataset was collected from Kaggle (California wildfire records, 2014–2025).

Total records: 700 (350 fire, 350 no-fire).

Main variables: temperature, NDVI, wind speed, relative humidity, road density.

Each variable was chosen based on its relevance to wildfire risk.

MODEL DEVELOPMENT



※ Four models were tested:

- ☑ Logistic Regression (LR)
- ☑ Support Vector Machine (SVM)
- ☑ Random Forest (RF)
- ☑ XGBoost

※ Each model was trained and compared using the same dataset.

※ XGBoost was chosen as the final model due to its higher accuracy and balanced performance.

※ Hyperparameters were adjusted through cross-validation to reduce overfitting.

RESULTS

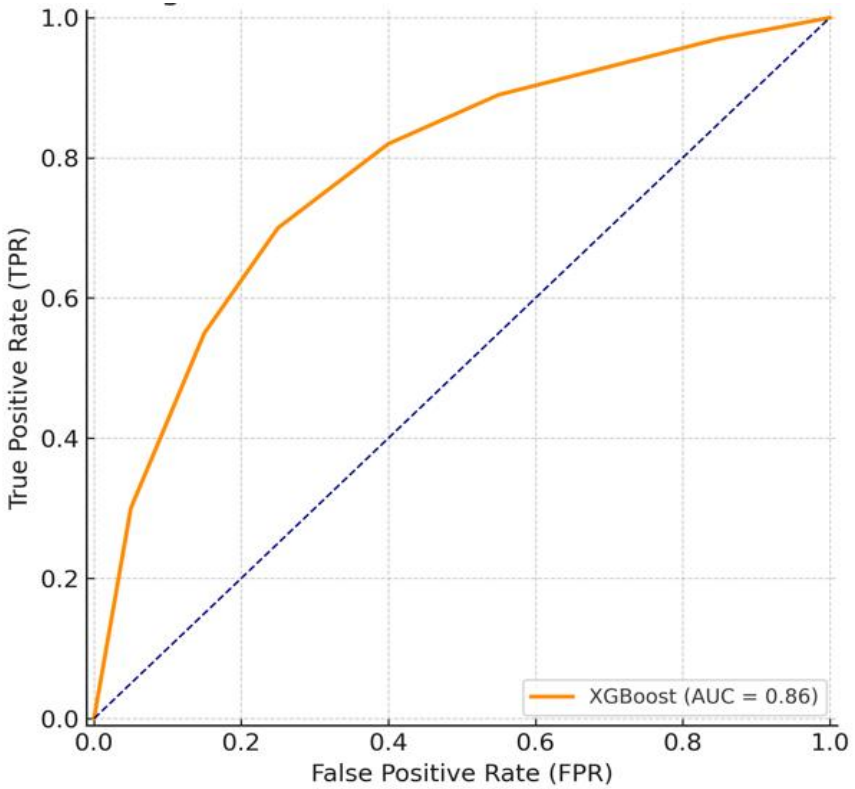
※ XGBoost showed the best performance among the four models.

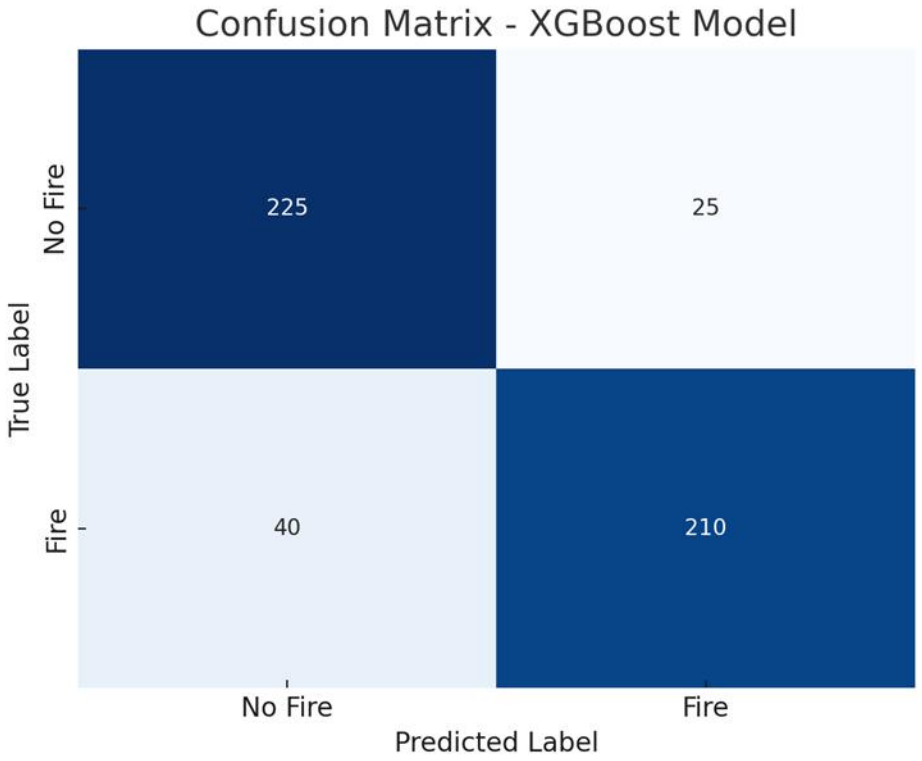
Overall Accuracy: 79.3% Precision: 89.4% Recall: 84.0%

F1-Score: 86.6% AUC: 0.86

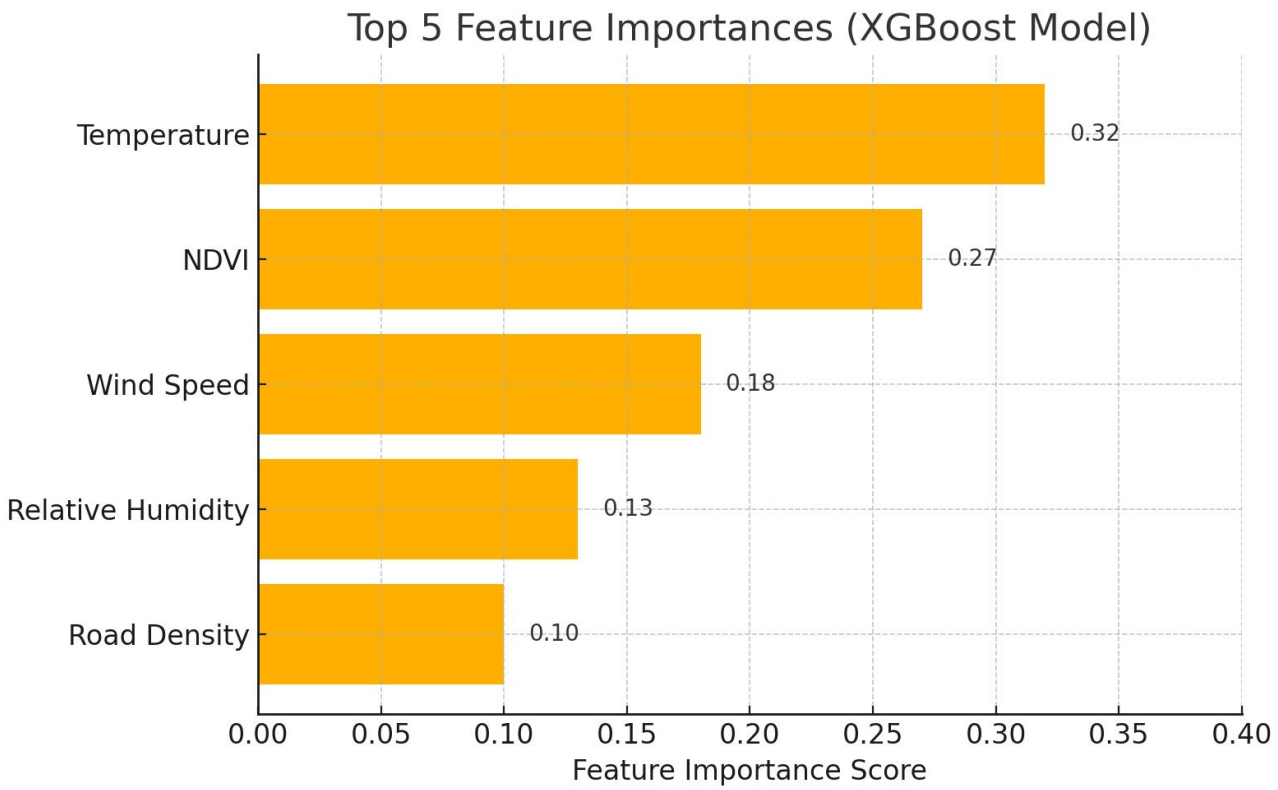
※Compared to Logistic Regression, SVM, and Random Forest, XGBoost balanced accuracy and stability better.

※5-fold CV mean accuracy: ~78% (±0.02)





DISCUSSION



※ The results show that XGBoost handled complex wildfire data better than other models.

※ Key factors like temperature, NDVI, and wind speed matched existing studies.

※ Some false positives and false negatives still appeared, mostly in regions with unusual weather.

※ Overall, the model aligns with practical expectations and literature benchmarks.

LIMITATIONS

⚠ The dataset was limited to California; results may not apply directly to other regions.

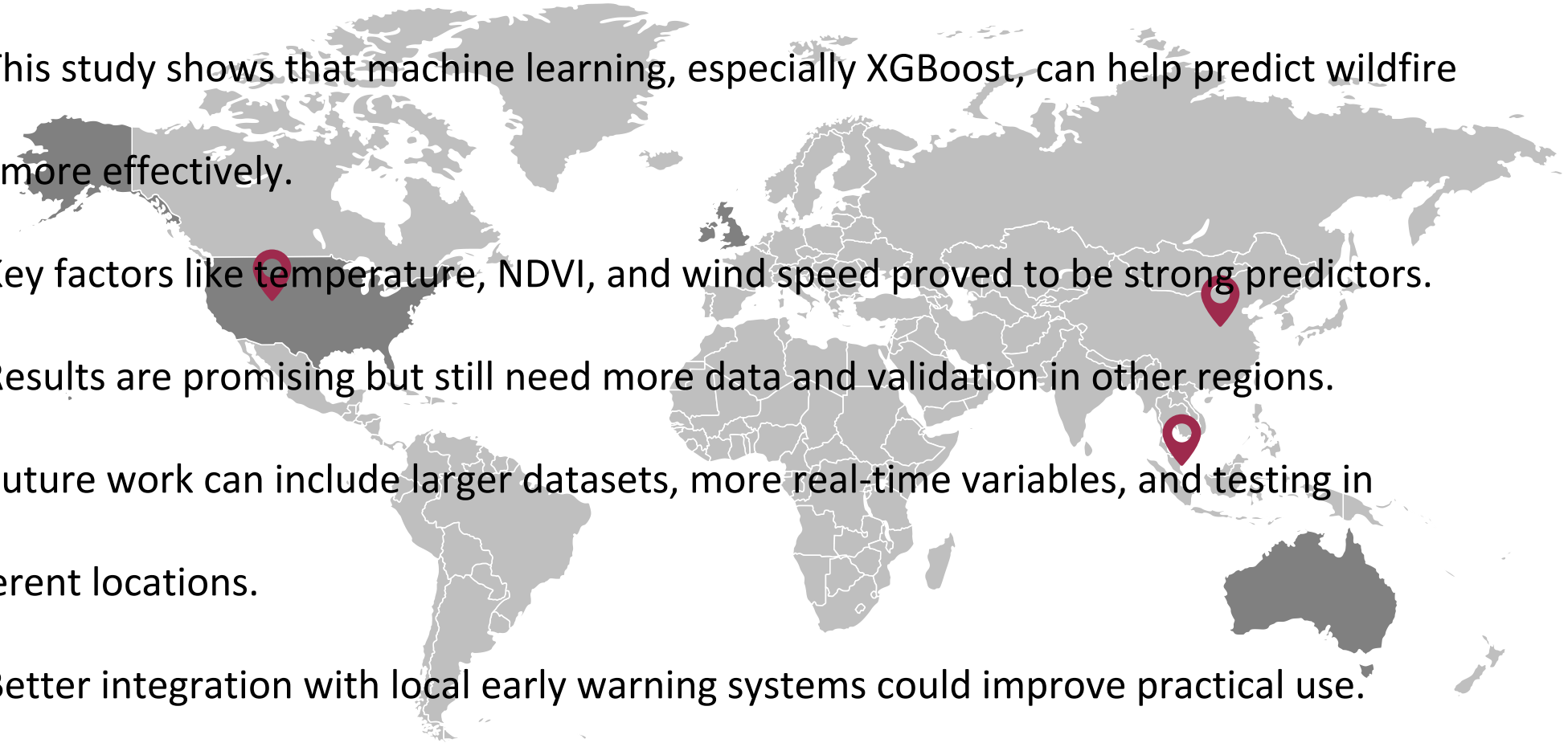
⚠ The total number of samples (700) is relatively small for some machine learning models.

⚠ Some data was missing or estimated, which may affect accuracy.

⚠ The model focused on specific variables; other potential factors like lightning or soil moisture were not included.

⚠ Predictions depend on the quality of real-time input data.

CONCLUSION & FUTURE WORK



※ This study shows that machine learning, especially XGBoost, can help predict wildfire risk more effectively.

※ Key factors like temperature, NDVI, and wind speed proved to be strong predictors.

※ Results are promising but still need more data and validation in other regions.

※ Future work can include larger datasets, more real-time variables, and testing in different locations.

※ Better integration with local early warning systems could improve practical use.

THANK YOU

univteknologimalaysia utm.my utmofficial