SCHOOL OF COMPUTING
Faculty of Engineering

Project Proposal Form MCST1043
Sem: 2 Session: 2024/25

SECTION A: Project Information.

Program Name:	**Masters of Science (Data Science)**
Subject Name:	**Project 1 (MCST1043)**
Student Name:	YI XINDIE
Metric Number:	MCS241027
Student Email & Phone:	Yixindie0219@outlook.com & 014 319 3841
Project Title:	Analysis of California's Decadal Wildfires: The Construction of Risk Management Models
Supervisor 1:	
Supervisor 2 / Industry Advisor(if any):	

SECTION B: Project Proposal

Introduction:

As a global hotspot for wildfires, California in the United States has significant reference value for wildfire analysis management measures worldwide. The frequent fires there are closely related to global climate change, human activities, and insufficient government control.

Therefore, we conducted a research and analysis on this issue. This is based on the publicly available wildfire data of the US government over the past decade, combined with data mining and machine learning techniques, and comprehensively considers the influences of meteorological conditions, geographical environment, and socio-economic factors to analyze the spatiotemporal distribution characteristics and risk driving mechanisms of wildfires. The research found that climatic conditions such as drought indices and Santa Ana winds are the main direct factors causing wildfires. At the same time, environmental changes caused by human activities have weakened the resilience of natural ecosystems to some extent, further increasing the probability of wildfires.

Based on the above results, this study constructed a wildfire risk prediction model based on data analysis and proposed the idea of social collaborative governance and policy recommendations for addressing climate change, aiming to provide certain references for related work.

Problem Background:

In recent years, with the intensification of global climate change and the continuous expansion of human activities, the frequency of extreme weather events has increased. Wildfires are not only an important natural

disaster problem faced by the United States, but also by all regions of the world. Based on the wildfire data released by the US government, this paper selects California as a typical case for analysis. Due to its unique geography, long-term high temperature and drought, and monsoon climate conditions, coupled with human activities such as population growth, urban expansion, and land pollution, the risk level of wildfires in California has continued to rise, and it has eventually become one of the most frequent wildfire areas in the world, bringing huge ecological damage and hundreds of millions of economic losses to California.

The frequent occurrence of wildfires not only destroys the local ecosystem, but also poses a certain threat to the entire earth's ecology. Therefore, how to scientifically control wildfire risks, predict in advance, and take effective prevention and response measures should not only be a problem that the United States should pay attention to, but also the whole world. In this context, data science provides new technical means for wildfire risk control and analysis. By integrating historical fire records, meteorological data, geological information and other data, and then using data mining and machine learning, it can more deeply reveal the spatiotemporal distribution characteristics of wildfires, and provide strong support for risk control and policy making.

Problem Statement:

In the past decade, the frequent wildfires in California have caused great damage to the regional and global ecology and huge economic losses, and have posed a continuous threat to the lives and safety of local residents. Although relevant departments have made some efforts in wildfire control in recent years and have achieved certain results, they still face many challenges in some aspects, which deserve further attention and research.

First of all, there are still new research possibilities for the existing wildfire risk prediction models in dealing with climate change and environmental changes caused by human activities, and there is still room for improvement in the accuracy and practicality of the prediction results.

addition, some high-risk areas themselves have strong risks. For example, in rural areas or areas with insufficient infrastructure, there is a lack of sufficient emergency equipment and adaptation personnel, which undoubtedly greatly increases the potential risk when wildfires come.

In response to the above problems, this study hopes to construct a risk prediction model based on data analysis to try to predict wildfire risks more effectively and provide certain reference suggestions. The model combines machine learning methods, as well as certain geospatial analysis techniques and socioeconomic factors to improve the accuracy and reference rate of predictions. It is also hoped that the model can provide certain support and optimization for resource allocation such as emergency response by using real-time data, and provide some valuable suggestions for wildfire control measures.

Overall, this study not only hopes to provide a certain reference for California, but also hopes that California, as a typical case, can provide some inspiration for related research in other wildfire-prone areas around the world.

Aim of the Project:

1. Improve wildfire risk forecasting by analyzing climate, terrain, and human factors

2. Optimize emergency response through better resource allocation

3. Support sustainable policies for long-term fire resilience

Objectives of the Project:
1. Using climate, topography, and human activity data to build more accurate wildfire prediction models.

2. Identify high-risk areas to provide a basis for improving emergency response.

3. Build a scalable analytical framework to provide reference for other fire-prone areas.

Scopes of the Project:
1. Geographic Focus: California or other wildfire-prone regions

2. Time Frame: Analysis of wildfire data from the past decade (2013–2023)

3. Technical Scope:

- Machine learning for risk prediction (e.g., Random Forest, LSTM)

- Geospatial analysis of fire-prone zones

Expected Contribution of the Project:
1. Accurate prediction: Combine multi-source data to improve early identification of wildfires.

2. Optimize response: Identify high-risk areas and improve resource scheduling and evacuation strategies.

3. Framework promotion: Build a transferable method that can be applied to other fire-prone areas.

Project Requirements:

Software: NumPy, Kaggle, Spark/Hadoop, Jupyter Notebooks

Hardware: AWS

Technology/Technique/ Methodology/Algorithm: Core Date Science& Machine Learning, Geospatial& Remote Sensing, Big Data& Cloud Computing, Visualization& Reporting

Type of Project (Focusing on Data Science):

[√] Data Preparation and Modeling

[√] Data Analysis and Visualization

[ ] Business Intelligence and Analytics

[√] Machine Learning and Prediction

[ ] Data Science Application in Business Domain

Status of Project:

[√] New

[ ] Continued

If continued, what is the previous title?

SECTION C: Declaration

I declare that this project is proposed by:

[✓] Myself

[ ] Supervisor/Industry Advisor ( )

Student Name: YI XINDIE

Signature **Date**

SECTION D: Supervisor Acknowledgement

The Supervisor(s) shall complete this section.

I/We agree to become the supervisor(s) for this student under aforesaid proposed title.

Name of Supervisor 1:

Signature **Date**

Name of Supervisor 2 (if any):

Signature **Date**

SECTION E: Evaluation Panel Approval

The Evaluator(s) shall complete this section.

Result:

[ ] FULL APPROVAL [ ] CONDITIONAL APPROVAL (Major)*

[ ] CONDITIONAL APPROVAL (Minor) [ ] FAIL*

* Student has to submit new proposal form considering the evaluators' comments.

Comments:

Name of Evaluator 1: .....

Signature ..... **Date** .....

Name of Evaluator 2: .....

Signature ..... **Date** .....