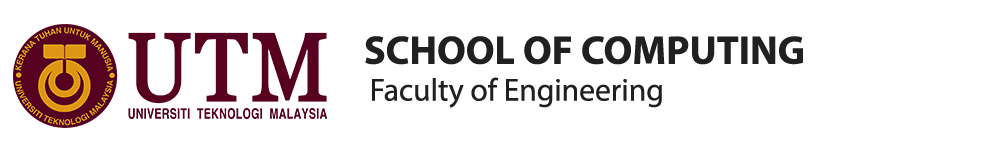
******

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](mailto:Yixindie0219@outlook.com) & 014 319 3841 | | |
| Project Title: | | An Analysis of California's Decadal Wildfires: From Data Mining to the Construction of | | |
|  | | Risk Management Models | | |
| Supervisor 1: | |  | | |
| Supervisor 2 / Industry Advisor(if any): | |  | | |

**SECTION B: Project Proposal**

**Introduction**:

|  |
| --- |
| As a typical sample of global wildfire management, California's frequent fire disasters are closely related to |
| climate change, the expansion of human activities, and deficiencies in management mechanisms. This research | |
| integrates meteorological, geographical, and socio-economic information based on a decade of wildfire data in | |
| California, and systematically analyzes the spatio-temporal distribution patterns of fire, risk driving mechanisms, | |
| and their socio-economic chain effects through data mining and machine learning models(such as random forests | |
| and LSTM). The study finds that drought index and climate factors such as the Santa Ana winds are core | |
| inducements for the spread of wildfires, while urban expansion exacerbates the vulnerability of high-risk areas. | |
| Additionally, the surge in insurance costs and post-disaster reconstruction pressures expose the limitations of | |
| traditional disaster prevention models. Based on this, the research proposes a data-driven wildfire risk prediction | |
| framework, community collaborative governance strategies, and climate adaptability policy recommendations, | |
| aiming to provide technical tools and decision support for interdisciplinary disaster management and promote | |
| resilient city construction under sustainable development goals. | |

**Problem Background**:

|  |
| --- |
| 1. High-Risk Area Expansion |
| According to the Cal Fire Department's March 2025 Fire Hazard Level Map, Southern California's "very high" |
| hazard area has grown 26 percent since 2011 to 817,000 acres, with about 4.6 million acres of land across the state |
| classified as medium, high, and very high. Chinese areas such as the area from Hasenda Gang to Chino Gang are |
| classified as the highest risk areas due to their proximity to mountainous vegetation areas, while plain areas such as |
| San Gabriel are at lower risk. |
| 2. Seasonal Patterns and Extreme Events |
| The peak period of fire occurrence is concentrated in the "Santa Ana wind" season from October to January, when |
| dry and strong winds (over 160 kilometers per hour) accelerate the spread of fires. The January 2025 "Palisades" and |
| "Eaton Fires" caused dozens of casualties, more than 10,000 buildings and billions of dollars in economic losses. |
| 3. Historical Disaster Comparison |
| - 2018-2020: Record-breaking fires, such as the 2018 Mendocino complex (1,858 square kilometers burned) and the |
| August 2020 complex fire (over 1 million acres), were caused by extreme drought and strong winds. |
| - 2021-2025: Warming climate exacerbates droughts, with 10,402 square kilometers burned in 2021, and the 2025 |
| Silver Fire threatens heritage and endangered species due to uncontrolled terrain and strong winds. ... |

**Problem Statement**:

|  |
| --- |
| Wildfires in California have become increasingly frequent and severe over the past decade, causing significant |
| ecological damage, economic losses, and threats to human safety. Despite advances in wildfire monitoring and |
| response systems, several critical challenges persist: |
| 1. Inadequate Predictive Models: Existing wildfire risk assessment methods often fail to account for dynamic |
| interactions between climate variables (e.g., drought, Santa Ana winds), human activities (e.g., urban expansion), |
| and environmental factors (e.g., vegetation density), leading to inaccurate predictions. |
| 2. Data Fragmentation: Wildfire-related data—including historical fire records, meteorological conditions, and |
| socioeconomic impacts—are scattered across multiple sources, making comprehensive analysis difficult. |
| 3. Limited Policy Effectiveness: Current disaster management strategies, such as firefighting resource allocation |
| and insurance pricing models, struggle to adapt to rapidly changing risk patterns, resulting in inefficient mitigation |
| efforts and rising financial burdens. |
| 4. Community Vulnerability: High-risk communities, particularly in suburban-wildland interface zones, lack |
| data-driven evacuation plans and resilience-building measures, increasing exposure to catastrophic fires. This study |
| aims to address these gaps by developing an integrated data science framework that combines machine learning, |
| geospatial analysis, and socioeconomic modeling to: |
| - Improve wildfire risk prediction accuracy through multi-factor dynamic modeling. |
| - Optimize emergency response and resource allocation using real-time data. |
| - Propose actionable policy recommendations for climate-adaptive wildfire management. |
| By bridging these research and application gaps, this work seeks to enhance California’s long-term resilience against |
| escalating wildfire threats while providing a scalable model for other fire-prone regions globally. |

**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. Reduce economic losses with smarter insurance and prevention strategies | |
| 4. Support sustainable policies for long-term fire resilience | |
| The results will help protect communities and guide wildfire management in fire-prone areas globally. | |

**Objectives of the Project**:

|  |
| --- |
| 1. Develop accurate wildfire prediction models using climate, terrain, and human activity data |
| 2. Identify high-risk zones to improve emergency response planning |
| 3. Analyze economic impacts to optimize insurance and recovery strategies |
| 4. Recommend sustainable policies for long-term wildfire prevention |
| 5. Create a scalable framework applicable to other fire-prone regions |
| These objectives aim to enhance California's wildfire resilience while providing solutions for global fire |
| management. |

**Scopes of the Project**:

|  |
| --- |
| 1. Geographic Focus: California (with potential scalability to 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 |
| - Socioeconomic impact assessment |
| 4. Limitations: |
| - Excludes real-time firefighting operations |
| - Focuses on pre-disaster risk reduction, not post-disaster recovery |
| - Relies on publicly available datasets |
| This scope ensures a focused yet impactful study on data-driven wildfire prevention and preparedness. |

**Expected Contribution of the Project**:

|  |
| --- |
| 1. Improved Wildfire Prediction– Develop a more accurate machine learning model for early risk detection by |
| integrating climate, terrain, and human factors. |
| 2. Data-Driven Emergency Planning – Identify high-risk zones to optimize firefighting resource allocation and |
| evacuation strategies. |
| 3. Economic & Policy Insights – Provide actionable recommendations for insurance adjustments and sustainable |
| wildfire prevention policies. |
| 4. Scalable Framework – Create a transferable methodology that can be adapted to other fire-prone regions globally. |
| 5. Public Awareness – Generate clear visualizations and reports to enhance community preparedness and |
| policymaker decision-making. |

**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** |