Project Proposal: Wildfire Risk Prediction Based on Temperature and Climate Change Scenarios

Wildfires pose increasing threats to ecosystems, human safety, and the economy. Climate change, through rising global temperatures and altered precipitation patterns, is a key driver of changing wildfire dynamics. This project aims to model the relationship between temperature and wildfire occurrence in the United States and to project how this risk may change under future warming scenarios derived from CMIP6 climate models.

Research Questions

1. How strongly is wildfire occurrence in the U.S. correlated with local temperature variations?
2. How does rising temperature affect the frequency and distribution of wildfires?
3. How will wildfire risk evolve under projected CMIP6 climate scenarios (e.g., SSP2-4.5, SSP5-8.5)?

Data Sources

1. **1.88 Million US Wildfires** (Kaggle, Rachael Tatman) — historical wildfire occurrences in the U.S. (1992–2015).
2. **Berkeley Earth Surface Temperature Data** (Kaggle) — monthly mean surface temperatures per city/region (1750–2015).
3. **WorldClim CMIP6 Projections** — projected future temperature data (2040–2100) under various SSP scenarios.

Methodology

1. **Data Preparation:** Join wildfire occurrence data with monthly temperature observations by location and time period. Clean and aggregate data at the state or county level.
2. **Feature Engineering:** Compute temperature anomalies, long-term trends, and seasonal averages. Include lag variables to capture delayed effects of heat on fire ignition.
3. **Modeling:** Train a classification model (e.g., Logistic Regression, LightGBM) to estimate the probability of a wildfire given temperature conditions.
4. **Projection:** Replace current temperatures with CMIP6-projected temperatures to estimate future wildfire probabilities and map the spatial change in risk ($\Delta p = p_{future} - p_{current}$).
5. **Evaluation:** Validate model performance with PR-AUC, Brier Score, and calibration metrics.

Expected Results

- Quantification of the relationship between temperature and wildfire occurrence.

- Spatial maps showing the projected increase in wildfire risk under +2°C and +4°C warming scenarios.
- Identification of U.S. regions most vulnerable to rising wildfire frequency.
- Policy-relevant insights for insurers, forestry agencies, and disaster risk management.

Dataset Overview

Table 1: Summary of datasets used in the project.

Dataset	**Variables**	**Period**	**Purpose**
US Wildfires	Location, Date, Fire Size	1992–2015	Target variable (fire occurrence)
Berkeley Earth	Temperature (monthly)	1750–2015	Historical predictor variable
CMIP6	Projected temperature	2040–2100	Future climate scenarios

Impact

Understanding temperature-driven wildfire risk dynamics is crucial for insurance and policy design. Results can inform premium adjustments, resource allocation for fire prevention, and long-term risk modeling under climate change. The simplified temperature-based model also serves as a scalable framework that can be expanded to include other environmental drivers such as precipitation and vegetation in future research.

Tools and Technologies

- Python (`pandas`, `scikit-learn`, `lightgbm`)
- `GeoPandas` / `Rasterio` for spatial joins
- `Matplotlib` / `Seaborn` for visualization
- Google Colab or local Jupyter environment for computation

Conclusion

This project will use publicly available datasets to model and visualize how increasing temperature affects wildfire probability in the U.S., and how this risk evolves under climate change. The focus on temperature as the main driver ensures transparency and interpretability, while integration with CMIP6 projections enables meaningful future scenario analysis.