Spring 2021

Wildfire Risk Prediction for a Smart City

Rekha Rani

San Jose State University

Wildfire Risk Prediction for a Smart City

A Project

Presented to

The Faculty of the Department of Computer Science

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Rekha Rani

May 2021

The Designated Project Committee Approves the Project Titled

Wildfire Risk Prediction for a Smart City

by

Rekha Rani

APPROVED FOR THE DEPARTMENT OF COMPUTER SCIENCE

SAN JOSÉ STATE UNIVERSITY

May 2021

Dr. Katerina Potika	Department of Computer Science
Dr. Navrati Saxena	Department of Computer Science
Dr. Ali Tohidi	Department of Mechanical Engineering

ABSTRACT

Wildfire Risk Prediction for a Smart City

by Rekha Rani

Wildfires are uncontrolled fires that may lead to the destruction of biodiversity, soil fertility, and human resources. There is a need for timely detection and prediction of wildfires to minimize their disastrous effects. In this research, we propose a wildfire prediction model that relies on multi-criteria decision making (MCDM) to explicitly evaluates multiple conflicting criteria in decision making and weave the wildfire risks into the city's resiliency plan. We incorporate fuzzy set theory to handle imprecision and uncertainties. In the process, we create a new data set that includes California cities' weather, vegetation, topography, and population density records. The model ranks the cities of California based on their risk of wildfires.

ACKNOWLEDGMENTS

I want to thank my project advisor, Dr. Katerina Potika, for her invaluable support and expertise in formulating the research questions and methodology. Her insightful feedback pushed me to work harder and brought my work to a higher level. Finally, I would like to thank my parents and my husband for their invaluable support throughout my life.

TABLE OF CONTENTS

CHAPTER

1 Introduction 1

1.1 Motivation 2

1.2 Problem Formulation 2

2 Definitions and Techniques 3

2.1 Wildfire Risk 3

2.2 Wildfire Risk Prediction 3

2.3 Smart City 3

2.4 Multi-Criteria Decision-Making 4

2.5 Fuzzy Set Theory 5

2.6 Data Mining 6

3 Related Work 8

3.1 Fire Weather Prediction 8

3.2 Fire Occurrence Prediction 11

3.3 Fire Risk Assessment 12

4 Methodology 14

4.1 Overview of the Data-set 15

4.2 Data Analysis 18

4.3 Wildfire risk prediction model 23

4.3.1 Weight determination 25

4.3.2 Calculation of global weighs for each criteria 30

4.3.3 Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) 30

4.4 Integration with smart city's resiliency plan 33

5 Experimental Results 34

5.1 California cities 34

5.1.1 Weather data 35

5.1.2 Vegetation data 37

5.1.3 Population density 38

5.1.4 Combining the data 38

5.1.5 Matrix formation 39

5.2 AHP fuzzy weight calculation 39

5.3 Weighted normalized matrix 39

5.4 California cities wildfire risk ranking 39

5.5 Model Accuracy 40

5.5.1 Historical data 41

5.5.2 Current data 41

5.6 Integration with Smart City 43

6 Conclusion and Future Work 45

LIST OF REFERENCES 46

LIST OF TABLES

1 Data Sources 13

2 Data Sources 19

3 Small prototype of the MCDM's criteria evaluation 24

4 AHP fundamental scale 25

5 Pairwise comparison matrix 26

6 Fuzzy AHP scale 26

7 Pairwise comparison matrix 27

8 fuzzy geometric mean 28

9 Relative fuzzy weight table 29

10 Defuzzification 30

11 Normalization 30

LIST OF FIGURES

1	Multi-criteria decision making process	4
2	Graphical representation of a crisp set	5
3	Graphical representation of a fuzzy set	6
4	Knowledge discovery of data	7
5	Weather factors for calculation of wildfire danger indices	10
6	Architecture of multi-layer perceptron	12
7	Implementation plan	14
8	Vegetation types in California	17
9	Data organization for quantifying fire risk	19
10	Fire Causes	20
11	Large Fire Causes	21
12	Fire Frequency	22
13	Fire Frequency	22
14	Correlation Test	23
15	Flow chart of prototype model	24
16	Calculation of global weights of criteria	31
17	Euclidean distance	33
18	California boundaries map	34
19	Merged dataframe	35
20	JSON weather data	36
21	Temperature and pressure from weather forecast	37

22 Processed API weather data 37
23 California county population map 38
24 Snapshot of data framework 38
25 Global weights calculation 39
26 Weighted normalized matrix of California cities 40
27 California cities wildfire risk ranking for smaller dataset 40
28 California cities from four counties 41
29 County map 42
30 Colby fire 2014 location, source:Wikipedia 42
31 Current wildfire season in California, source:Wikipedia 43
32 Wildfire risk in California 43
33 Fire hazard severity zone map, Source: CA.gov 44
34 Alert email snapshot 44

CHAPTER 1

Introduction

Forest fires are considered a constructive force of nature as they shape the ecosystem and help in the replenishment of wood. These beneficial fires create open spaces that allow young trees to get enough sunlight and nutrients. However, a forest fire can become a destructive force when it damages properties and claims human lives. Uncontrolled wildfires also contribute to the degradation of air quality. Moreover, an unchecked fire burning at a high temperature is termed as wildfire as it may destroy forests and organic matter. Therefore, timely detection and prediction of wildfires are required to minimize their disastrous effects.

Unfortunately, wildfire prediction is a challenging task. It is difficult to develop an accurate early-warning system because random human actions ignite many wildfires. Nevertheless, forecasters use various factors like climatic conditions, vegetation types, etc., to issue warnings for naturally occurring wildfires. Several existing technologies have been proposed, and various models have been implemented that detect fires and wildfires, e.g., wireless sensor networks, feed-forward artificial neural networks (ANN) [1], fire weather prediction using self-organizing maps [2]. The recent research on wildfire risks by Professor Jerry Gao [3] at San Jose State University has predicted the wildfire risks at Monticello and Winters in California using random forest models.

In this research, we propose a wildfire prediction model that relies on the multi-criteria decision-making (MCDM) technique to explicitly evaluate multiple conflicting criteria in decision making and weave the wildfire risks into the city's resiliency plan. We used fuzzy sets to handle imprecision and uncertainties and performed several experiments using the historical wildfire data to check the accuracy of our model. Additionally, we combined multiple data sets and emphasized ranking the areas in accordance with the fire risks.

1.1 Motivation

The year 2020 saw California's largest wildfire season with 367 known fires. These concurrent wildfires damaged nearly 100 million acres of land, and more than 60,000 people were forced to evacuate [4]. The Colossal smoke clouds aggravated air quality, and a state of emergency was declared. This horrific scenario presented an opportunity to understand better how the wildfire spread and how to predict it. Therefore, we decided to build a wildfire risk prediction model to help the city plan for high-risk zones.

1.2 Problem Formulation

The wildfire risk prediction problem presented in this research is solved in the following phases:

1. Data Integration from various sources

   Wildfire depends on multiple factors, so we integrated various types of data that would provide details about vegetation, climate, population density, latitude, longitude, and slope of an area.

2. Exploratory analysis of the data

   We performed exploratory analysis on the prepared data set and observed existing patterns in the wildfires throughout California.

3. Wildfire risk prediction model

   After observing wildfires' patterns, a wildfire risk prediction model is developed using MCDM to rank different areas according to fire risks.

4. Weaving wildfire risk into the city's resiliency plan

   We integrated the fire risk into the city's infrastructure plan and resiliency plan for high-risk zone area.

CHAPTER 2

Definitions and Techniques

This chapter will define the techniques and strategies that are used in our research.

2.1 Wildfire Risk

S. Kaplan and B.J. Garrick [5] define the term risk as the possibility of an unfortunate occurrence. It is the probability of happening something harmful or undesirable.

$$Risk = Uncertainty + Damage$$

In the wildfire context, we have to modify this definition since the fire risk can bring either uncertain damage or uncertain benefit depending on whether the fire is a wildfire or beneficial fire. So, wildfire risk can be defined as the combination of the probability of wildfire, the intensity of a wildfire, and the effects of wildfire.

2.2 Wildfire Risk Prediction

Predicting is the process of forecasting what might happen. We have to consider a range of possible outcomes to predict the future. Since unplanned wildfires can impact the ecological and social systems, there is a need to anticipate future fires. Additionally, it is impractical to maintain the firefighting units active in all parts of the city. Hence, a city needs to assess the wildfire risk in advance to incorporate the wildfire risks of an area into its resiliency plan [6].

2.3 Smart City

The smart city is the concept that supports modernizing urban life using robust strategies and innovative planning [7]. In a smart city, the city policymakers use information technology to deliver services efficiently and sustainably. The research would integrate the wildfire risks and ranking of an area into a smart city's planning model. By knowing the fire risks of a site, a city can build multiple fire stations in critical zones. City authorities can incorporate information technology to send fire alerts to the residents.

2.4 Multi-Criteria Decision-Making

Decision-making is a cognitive process of making a choice based on various assumptions, preferences, and several other factors that result in selecting a belief or a plan of action among several probable options [8]. During a decision-making process, a large amount of data or information need to be processed to reach a rational decision. Such information may be incomplete, inconsistent, and conflicting with one another. So, decision-making with traditional methods may not be fruitful in these scenarios. The multi-criteria decision making (MCDM) technique improves the quality of decisions by considering several criteria and alternatives in a more efficient and rational way. MCDM is widely used for decision-making in various fields like business, economy, disaster management, etc. [9]. Figure 1 explains the process of decision making, using a multi-criteria decision-making technique. Figure 1.

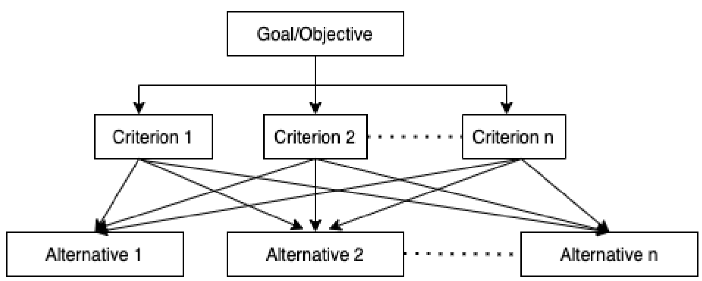


Figure 1: Multi-criteria decision making process

Our proposed model considers the weather, topography, and vegetation of an area as they play a significant role in predicting wildfires. Some other factors, like lower relative humidity, stronger winds, and hotter temperatures, increase wildfire chances, so we also incorporated them. Using population density data and ground data for a region, we included the human factors in our model.

2.5 Fuzzy Set Theory

A crisp or a classical set is an unordered collection of different elements with fixed and well-defined boundaries. It can be represented using a characteristic function as explained in Figure 2.

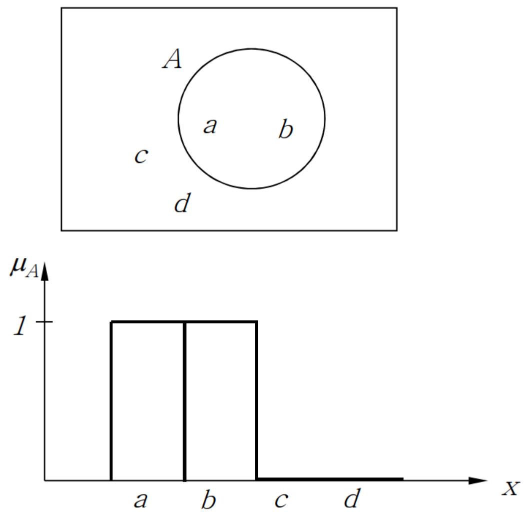


Figure 2: Graphical representation of a crisp set

$$A = a_1, a_2 \ldots, a_n$$

$$\begin{cases} \mu(x) = 1, if x \in A \\ \mu(x) = 0, if x \notin A \end{cases}$$

We cannot rely on classical set theory for real-world problems as real-world problems are associated with uncertainties and do not have well-defined boundaries. A fuzzy set is a set with imprecise or vague boundaries [10]. It is an extension of the classical set and a potential tool for handling imprecision and uncertainties. We used fuzzy sets in our project to find an approximate solution that handles imprecision and uncertainties. Figure 3 depicts graphical representation of a fuzzy set.

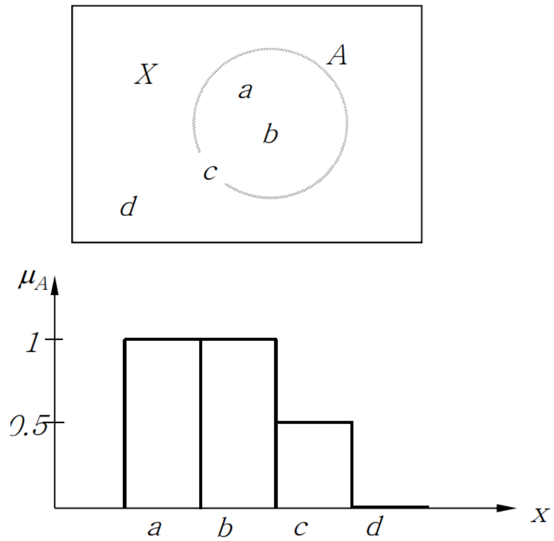


Figure 3: Graphical representation of a fuzzy set

$$A = a_1, a_2 \ldots, a_n$$

$$\begin{cases} \mu(x) = 1 \\ \mu(x) = 0.5 \\ \mu(x) = 0 \ldots \end{cases}$$

2.6 Data Mining

The data required for wildfire risk prediction is massive and growing rapidly. So, it is not possible to convert this huge amount of data into knowledge manually. Therefore, we relied on various data mining techniques to extract information and process the data. Data mining is also known as information harvesting or knowledge discovery as it provides various technologies to make knowledge-driven decisions[11]. Figure 4 represents various steps in the knowledge discovery of data that are summarized below.

- `Data Integration`: Combining data from various sources.
- `Data Cleaning`: Removing duplicate or irrelevant observations.

- `Data Selection`: Retrieving relevant data from the database.
- `Data Transformation`: Transforming data into an appropriate form suitable for data mining
- `Data Mining`: Transforming the data into patterns.
- `Pattern Evaluation`: Interpreting mined patterns using summarization and visualization.
- `Knowledge representation`: Generating reports, tables, discriminant rules, and classification rules.

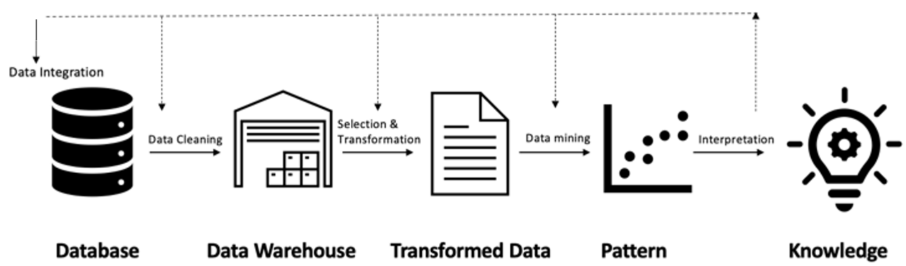


Figure 4: Knowledge discovery of data

CHAPTER 3

Related Work

Wildfire is a complex phenomenon; it is a product of several interrelated factors like weather and topography. While the complexities of wildfire present challenges, with the advancement of various machine learning and remote sensing techniques, notable improvement has been made in wildfire prediction. This chapter will explain different wildfire prediction techniques that will help us understand the problem domain. We can divide the wildfire problem into three main domains:

- Fire weather prediction
- Fire occurrence prediction
- Fire risk analysis

3.1 Fire Weather Prediction

Various studies have shown that weather conditions play a significant role in the start of wildfires, forest fuel combustibility, and wildfire behavior. Also, wildfire behavior is greatly affected by topography and fuel type. The data for these parameters can be obtained from the local meteorological department. European Forest Fire Information System (EFFIS) provides a framework to monitor and forecast fire danger in Europe using weather forecasts [12].

The Weather observations can also be used to calculate fire danger indexes, such as the National Fire-Danger Rating System (NFDRS) [13]. The main input components into the NFDRS model are vegetation fuels, weather, and topography. Similarly, the Canadian Fire Weather Index (FWI) [14], is a meteorologically based fire danger index that considers temperature, humidity, precipitation, wind, and moisture contents of fuel. FWI system relies upon the consistency of fuel to calculates the fuel moisture contents. Various fuel moisture codes like fine fuel moisture code, duff moisture code, etc., are used to provide numeric fire intensity ratings. These moisture codes are summarized below:

- `Fine Fuel Moisture Code (FFMC)` - Fine fuel moisture code is a numeric grading scale of the average moisture content of forest litter and fine fuels.
- `Duff Moisture Code (DMC)` - DMC is a numeric grading of the average moisture content of decomposed compacted organic material at moderate depth.
- `Drought Code (DC)` - DC is a numeric grading of moisture content of deep organic layers, and it indicates the effects of drought on forest fuel.

There are a few papers that address fire weather prediction using machine learning methods. We cannot rely on traditional statistical since the relationship between synoptic weather and wildfire is nonlinear and high-dimensional [15]. Since self-organizing maps (SOMs) have the ability to learn nonlinear relationships and handle high-dimensional data, so they are used in many meteorological studies. R. Lagerquist et al. [2] trained SOMs to predict the fire weather in northern Alberta. They produced various map types and associated different fire-weather climatology with each map type.

M. Crimmins [16] proposed a slightly different approach with the synoptic-pattern classification method to identify different synoptic weather patterns across the southwest USA. He used Fosberg Fire-Weather Index (FFWI), a nonlinear filter, to determine the fire danger levels. The value of FFWI increases with the increase in the speed of the wind and with the decrease in relative humidity. The SOM synoptic classification method produced three weather types that were associated with extreme surface fire-weather conditions.

Various researches have been conducted for lightning prediction. Some researchers used the lightning prediction model for wildfire prediction as lightning is one of the major causes of wildfires. M. Pakdaman et al. [17] proposed an ensemble algorithm for lightning prediction. This ensemble algorithm can be integrated with wildfire prediction models.

We can easily conclude that the weather plays a vital role in wildfires, and we should consider these weather factors for wildfire prediction models. Some of the important weather factors depicted in Figure 5 are explained below:

- `Temperature`: Higher temperatures create dry conditions and make the fuel more susceptible to wildfires.
- `Precipitation`: Higher precipitation can add more moisture to fuels; therefore, it can act as a negative indicator of wildfire spread. But, higher precipitation can increase the vegetation cover of an area, thereby increasing the fuel and increasing the chances of wildfires.
- `Evapotranspiration`: Evapotranspiration is the measure of both evaporation and plants' transpiration from the earth's surface to the atmosphere [18]. When the rate of evapotranspiration is higher, fire fuel is more prone to wildfires.
- `Wind`: Strong winds supply oxygen to fire and preheat the nearby fuels. Therefore, strong winds increase the spread area of wildfires.
- `Relative Humidity`: The relative humidity is a measure of the moisture contents in the atmosphere. The lower the relative humidity, the drier the atmosphere. And the drier the atmosphere, the more readily a fire will start.

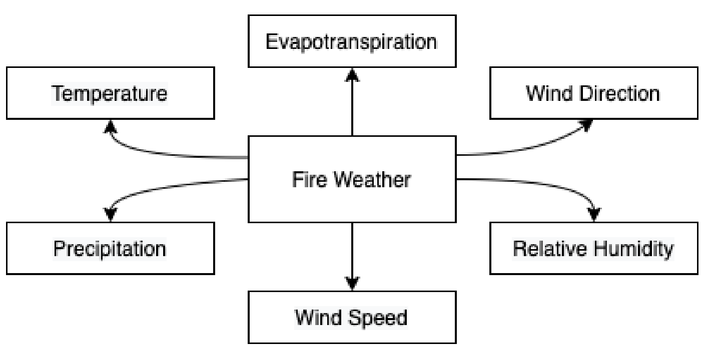


Figure 5: Weather factors for calculation of wildfire danger indices

So, we can say that lower relative humidity, stronger winds, and hotter temperatures increase wildfire chances. The data obtained from weather stations can be used to calculate the meteorological-based fire danger indexes. However, errors may occur in calculating fire danger indexes that would lead to false alarms for wildfire.

3.2 Fire Occurrence Prediction

Predicting future fires' occurrence plays a vital role in allocating resources, recovery efforts, and preparedness planning. Most researchers rely on Artificial neural networks (ANNs), a machine learning method, to predict fire occurrences. An artificial neural network (ANN) is a system designed to simulate the human brain's functioning to analyze and process information. Therefore, ANNs can quickly solve the problem that is rich in data, i.e., the issues with several examples to train the model.

Alonso-Betanzos et al. [1] used feed-forward artificial neural networks (ANN) to predict a fire occurrence risk index using the data obtained from five Galician meteorological stations. Similarly, Y. Safi and A. Bouroumi [19] used multi-layer perceptron to predict the wildfires. They used the back-propagation learning algorithm to train the ANNs since its optimization procedure minimizes the global error observed at the output layer. An architecture of multi-layer perceptron is depicted in Figure 6.

While ANNs based model easily solves the wildfire problem, the major shortcoming of these models is that they require substantial computational resources. Secondly, the neural networks are a "black box" and cannot easily identify casual relationships.

H. Naganathan et al.[20] used different predictive methods like K-Nearest, support vector machine, and decision tree to predict wildfire occurrences. They used the meteorological and fire data to check the accuracy of these models. We know that ANNs based methods were relying on a massive amount of data. Therefore, G.E. Sakr [21] proposed a model using a support vector machine to predict wildfire risk

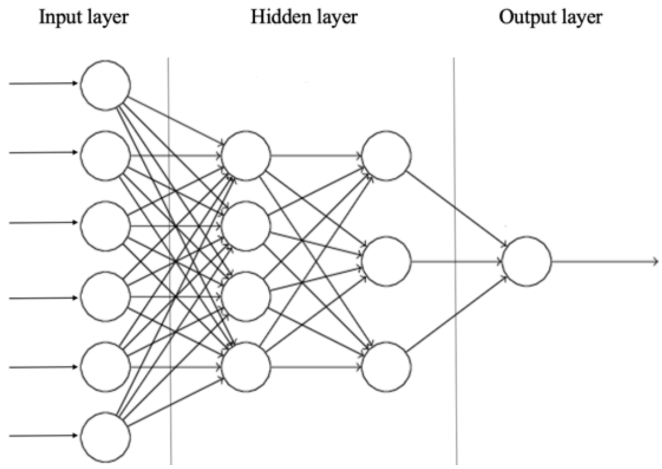


Figure 6: Architecture of multi-layer perceptron

with limited data. Their model did not rely on any weather prediction mechanism and used only the meteorological data. They introduced a fire risk index that corresponds to the possible number of fires on a particular day.

3.3 Fire Risk Assessment

Wildfires can severely impact our ecological, social, and economic systems. Therefore, there is a need to estimate and assess the risks posed by wildfires. Risk assessment can be considered as a decision support tool for strategic and tactical decision-making. Analyzing wildfire risks would help the city authorities to make decisions where consequences are intrinsically uncertain.

The resources to handle natural calamities are limited in number. Therefore, these resources should be allocated carefully after prioritizing the risk zones. Multi-criteria decision-making is a technique proposed to solve decision-making and planning problems that involve multiple criteria [9]. G. Jakovljević et al. [22] proposed a model based on Geographical Information Systems (GIS) and multi-criteria decision making. The model divided the area map of Municipality Nevesinje, Bosnia, and Herzegovina into five categories. They categorized the area map so that category one areas were at the lowest risks and areas with category five at the highest risks of wildfires.

A. Lapucci et al. [23] proposed a slightly different approach with MCDM. They used knowledge discovery of data along with spatial MCDM model for fire risk evaluation. The model identified the areas that are subject to higher fire probability. Although this model successfully evaluated the fire risks, there is still scope for improvement. We can improve these models using modified criteria and, fuzzy set theory, and sensitivity analysis.

Table 1 describes the data sources used in related works.

Table 1: Data Sources

Data Category	Source
Weather	The National Oceanic and Atmospheric Administration National Centers for Environmental Prediction wildfire.alberta.ca
Vegetation	berkeley.edu github data.mendeley.com/datasets/85t28npyv7/1
Fire History	UC Irvine Machine Learning Repository MONITORING TRENDS IN BURN SEVERITY (MTBS) Fire Weather Knowledge Centre, Australia Geoscience Austrialia data.mendeley.com/datasets/85t28npyv7/1 Kaggle

CHAPTER 4

Methodology

This project's fundamental goal is to develop a wildfire risk prediction model using MCDM methodology in conjunction with fuzzy logic. The proposed model ranks and prioritizes the different areas per the fire risks. The city can use the risk ranking for preparedness planning of risk zones. The implementation plan for our project is explained in Figure 7.

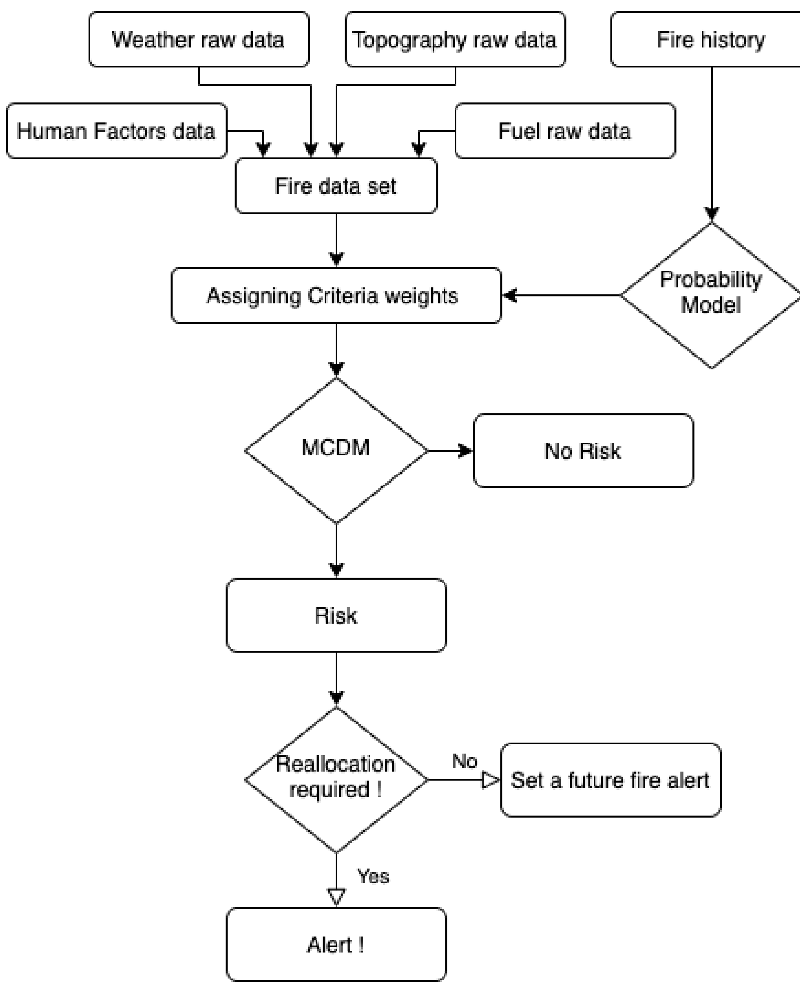


Figure 7: Implementation plan

4.1 Overview of the Data-set

The data-set used in the project includes fire history data, weather data, vegetation data, population data, topography data for evaluating the fire risk in the study areas. The data-set contains various sub-parameters of weather and vegetation data:

- Fire history data:

  The SQLite fire history data obtained from Kaggle contains the history of wildfires in the United States between 1992 and 2015. The dataset contains various files describing the wildfire history. We used the table Fire for our project. The Fire table data comprises different attributes like fire name, a global unique identifier for fires, source database, local report fire ID, etc. We extracted the following attributes for fire history records for the state of California.

  - FIRE CODE: Code used by wild land fire communities
  - FIRE NAME: Name of the fire
  - FIRE YEAR: Calendar year when the fire occurred
  - DISCOVERY DATE: Calendar day when the fire occurred
  - DISCOVERY DAY: Day of the year when the fire occurred
  - DISCOVERY TIME: Time when the fire occurred
  - STAT CAUSE CODE: Code describing the cause of wildfire
  - STAT CAUSE DESCR: Description for the cause of wildfire
  - FIRE SIZE: Acres of fire parameter
  - FIRE SIZE CLASS: fire size class depending on the Acres burnt.
  - LATITUDE: Latitude of the location of fire
  - LONGITUDE: Longitude of the location of fire
  - STATE: State where the fire occurred
  - COUNTY: County where the fire occurred.

We can obtain the probability of wildfires in these areas by studying the various parameters.

- Weather data: Fire occurrence and fire spread are dependent on the various climatic factors of a place. We used weather data from two sources:

  1) Weather data for fire history analysis:

  To analyze the impact of weather on the wildfire, we obtained the monthly data from the National Oceanic and Atmospheric Administration.

  2) Weather data for prediction:

  We used OpenWeatherMap API to retrieve the weather forecast for a given region. It returns the result in JSON format. We used the temperature, atmospheric pressure, humidity, and wind speed to forecast the chances of wildfires in a given region.

  We analyzed multiple factors responsible for wildfire:

  - Temperature: Temperature has a direct relationship with the dryness of fuel. The more the temperature, the drier the fuel. If the fuel is dry, it would catch the wildfire easily
  - Wind speed: The rate of spread of a wildfire increases with the increase in wind speed. Strong winds can also lead to spark in power lines, and this spark can be converted into wildfire if the nearby fuel is dry.
  - Relative humidity: Low relative humidity increases fire behavior because it makes the fuel drier.
  - Precipitation: Higher precipitation can increase the vegetation cover of an area. We know that vegetation acts as fuel for fires, so, more fuel can increase the chances of wildfire.

- Vegetation data: Vegetation is an essential factor in the prediction of a wildfire. We know that vegetation act as fuel for wildfires. There are different types of vegetation, and some are more combustible than others. The areas that have more fuel would have more chances of wildfires than the area that has less fuel. The percentage of fuel is directly proportional to the percentage of biomass at a place. The shapefile obtained from Wieslander's vegetation type mapping contains numerous types of vegetation in California. We mapped the vegetation on California's state boundaries which is depicted in Figure 8.

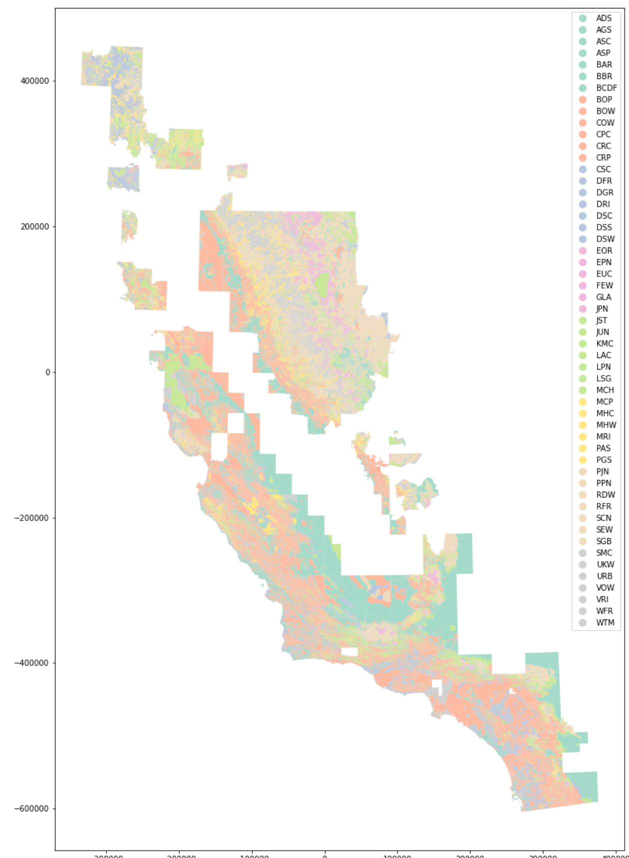


Figure 8: Vegetation types in California

These categories of vegetation are re-categorized into a few categorized depending on flammability. We will consider the deciduous forest percentage, grassland percentage, evergreen forest percentage, etc. category in our model.

- Topography: The geography of a place is a significant factor for the start of wildfire.
  - Slope: The slope of an area is important for fire risk prediction as the fire travels at a very high speed up-slope than at low-slope.
  - Location: The longitude and latitude of a place decide the circulatory wind pattern and the solar cycle.
- Human Factors Data: According to the U.S. Department of Interior, 88% of wildfires are caused by human negligence. So, human factors should be considered carefully.
  - Population: Number of people living in an area if obtained city-wise and county wise.
  - Population Density: Population Density is defined as population per unit area. If the population density of an area is more, the area would have more picnic spots and more fire camps than the areas with less population density.
  - Area of Land: Area of Land of a particular geographic region is obtained from US Census Bureau.
  - Area of Water: Area of Land of a particular geographic region is obtained from US Census Bureau.

4.2 Data Analysis

- Phase 1: Integration of data from multiple sources:

  In this phase, data from various sources like weather data, topography data, fuel data, and human factors data in our data-set. Table 2 describes the data sources for our project.
- Phase 2: Pre-processing of the data-set:

Table 2: Data Sources

Data Category	Source
Weather	The National Oceanic and Atmospheric Administration OpenWeatherMap
Vegetation	berkeley.edu github
Social Factors	Census Bureau Wikipedia
Fire History	Kaggle
Social Factors	Census Bureau Wikipedia
Geography	simplemaps.com

In this phase, we pre-processed the data. We need only a few parameters for our model so, we extracted only the necessary attributes from the data-set [24]. Various attributes that would be included in our data-set are described in Figure 9.

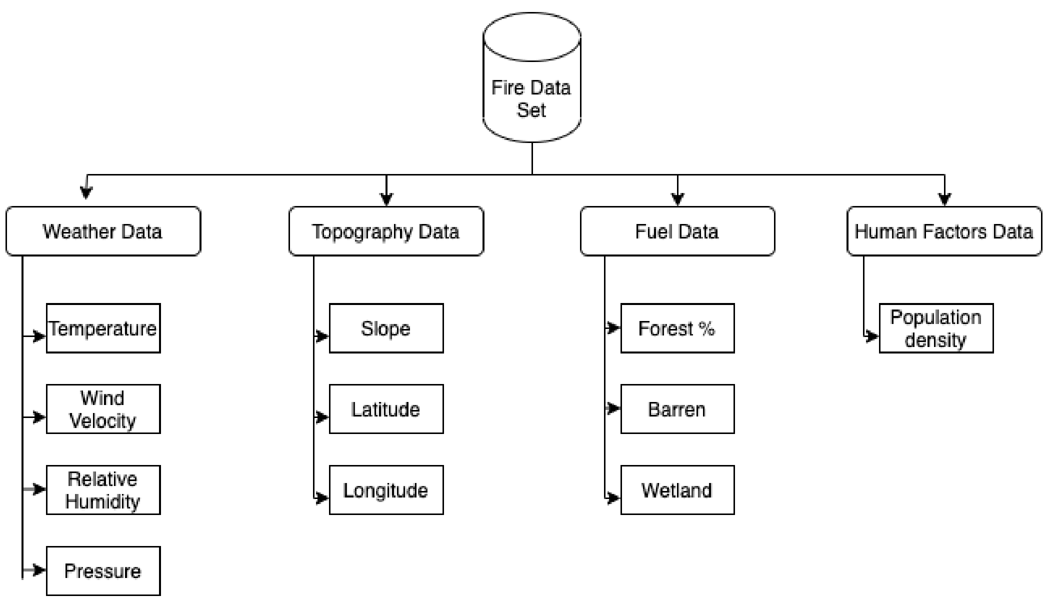


Figure 9: Data organization for quantifying fire risk

- Phase 3: Data factors analysis:
  - Cause of wildfire:

    We counted the number of wildfires and analyzed the STAT CAUSE DESCR column in the Kaggle fire history database for the state of California. Figure 10 explains that human factors and lightening are the major causes of wildfire in California over the past few years.

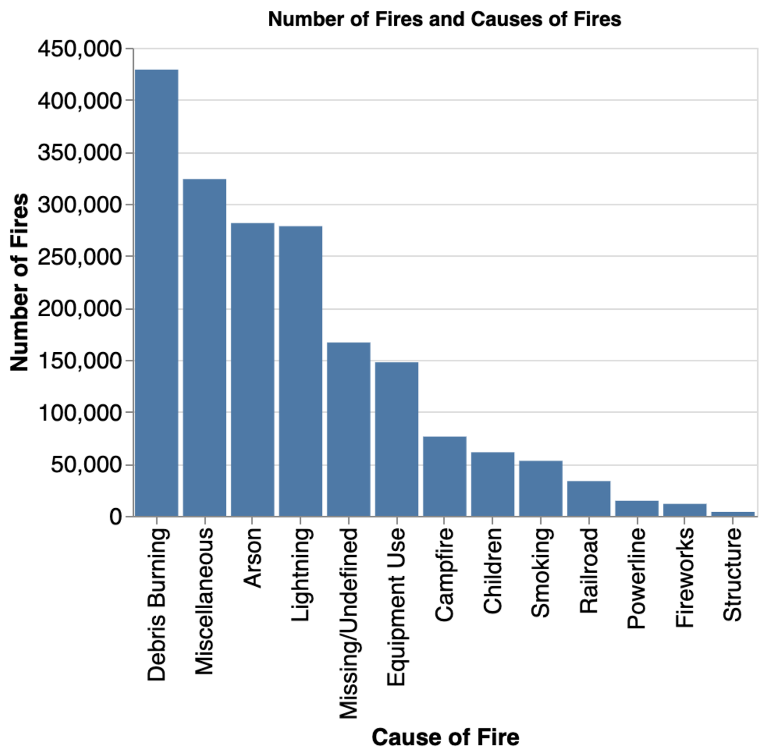


Figure 10: Fire Causes

After analyzing the causes of wildfire by considering only those fires that burned more than 500 acres of land, the primary cause of the large wildfire is found to be lightening as described in figure 11. So, weather is an important factor that should be consider for prediction model.

  - Monthly frequency of wildfire:

    The heat map in figure 12is obtained from fire history data-set. We can say

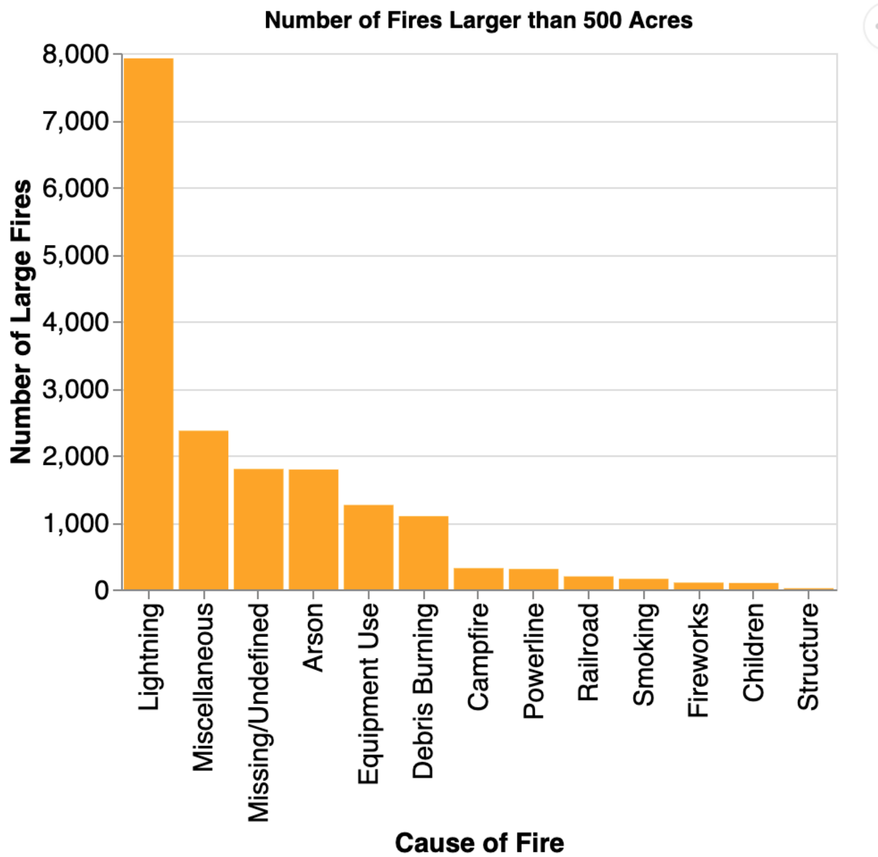


Figure 11: Large Fire Causes

that fires are more frequent in the months of June and July in California. The state of California lies in subtropical climate zone. June and July are months of summer season and high temperature. So, temperature of a region is considered an important factor for starting of the wildfires in our model.

– Fire count in each county:

  The topography, vegetation, and weather of a place play an essential role in wildfires. The fire count occurrences in some areas are more because of its geography and climate. The count of the large fires with fire size greater than 500 acres is depicted in figure 13

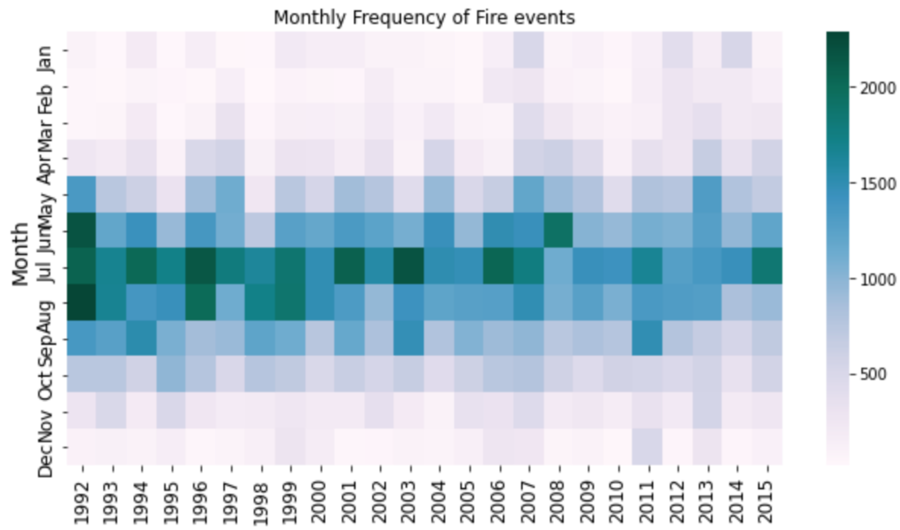


Figure 12: Fire Frequency

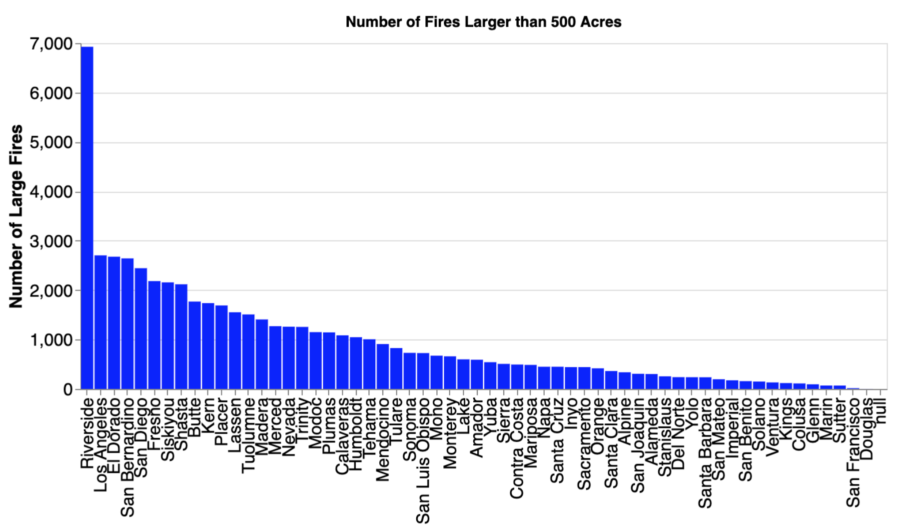


Figure 13: Fire Frequency

– Multicollinearity among numerical variables:

We performed the statistical correlation test to find whether there is a liner relationship between two quantitative variables. The result of the correlation test between fire count, area of land, deciduous forest percentage, evergreen forest percentage is depicted in Figure 14.

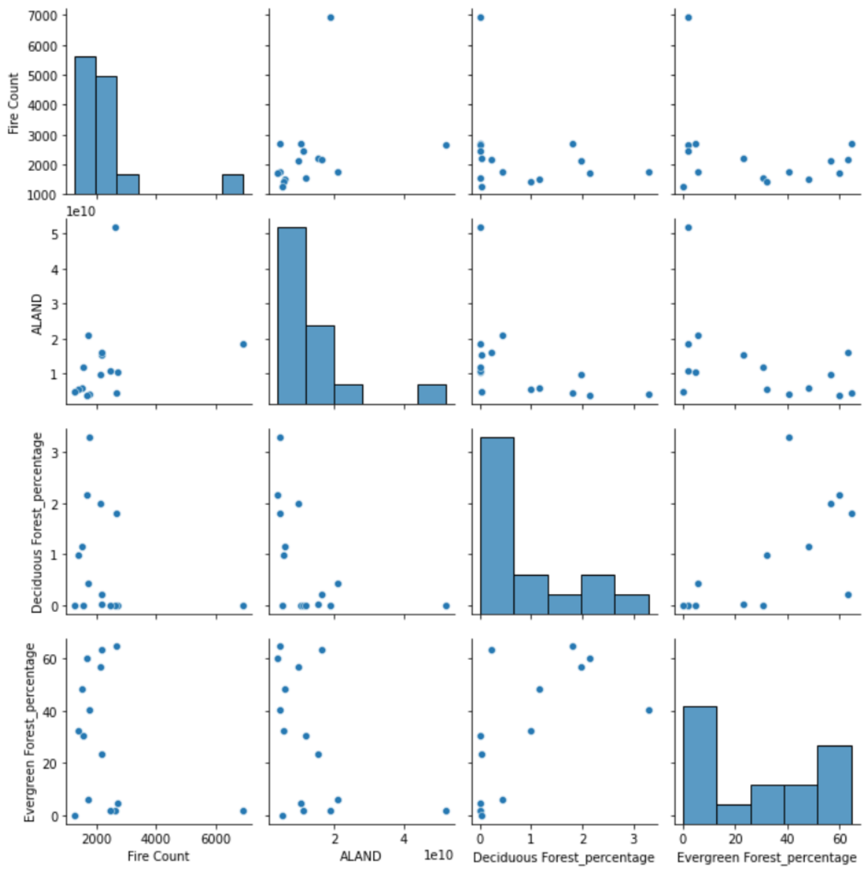


Figure 14: Correlation Test

4.3 Wildfire risk prediction model

We built a wildfire risk prediction model using multi-criteria decision banking to rank different areas according to fire risks. By analyzing the date, we found that each criterion is not equally important. So, we used a weighted matrix for different criteria. We relied on fuzzy set to handle imprecision and uncertainties since fuzzy sets use a degree-based membership function. To determine the weights of all the criteria, analytical hierarchy process is used. A prototype model explaining MCDM has been described in Table 3. Figure 15 below is the flow chart describing steps that should be followed to determine the ranking using MCDM.

- The prototype model relies on three factors, factor 1, factor2, and factor3 to

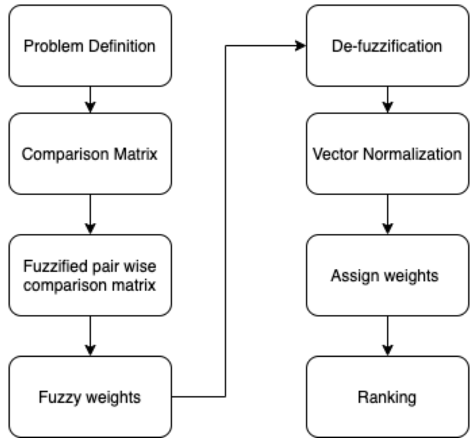


Figure 15: Flow chart of prototype model

predict the ranking of future wildfires in given regions.

- We will assign the weights to all the attributes of our model.
  - Area = ['A', 'B', 'C', 'D', 'E']
  - Attributes = [Factor1, Factor2, Factor3]
  - Weights = [0.2, 0.2, 0.6]
- After performing vector normalization and separation measurements, the output of this model provides a ranking of all the areas following fire risks. The last column in Table 3 depicts the result i.e., the order of all the regions.

Table 3: Small prototype of the MCDM's criteria evaluation

Area	Factor1	Factor2	Factor3	Rank
E	8.2	51	6.7	1
C	18	33	0.9	2
B	14.6	33	1.3	4
D	8.3	97	4	3
A	11.4	99	1.8	5

4.3.1 Weight determination

Our model relies on the analytic hierarchy process (AHP) to calculate the weights by pairwise comparisons. The factors are compared based on data analysis performed on California's wildfire history data from 1992-2015. We used Saaty's scale to prioritize the factors of wildfire in the given areas. This scale is used for pairwise comparisons by mapping the relative importance of different factors to value ranging between 0 and 9, as shown in Table 4.

Table 4: AHP fundamental scale

Intensity of Importance	Definition
1	Equal importance
3	Moderate importance of one over another
5	Essential or strong importance
7	Very strong importance
9	Extreme importance
2,4,6,8	Intermediate values

Criteria are compared with one another by forming a n*n matrix and prioritized using pair-wise comparisons. For example, if factor1 is equally important to factor2, then it is assigned a weight of 1. So, factor2 weight would become 1/1 with respect to factor 1. Similarly, if Factor1 is extremely important with respect to Factor3, it is assigned a weight of 9. So, Factor3 weight would become 1/3 with respect to Factor1.

4.3.1.1 Construction of FAHP comparison matrices

In real world problems, it is difficult to map qualitative preferences to crisp values. So, we will use a fuzzy scale of relative importance based on Saaty that is described in table 6 to create a pairwise comparison matrix table 5.

- Table 4 is used to convert the point preferences to triangular fuzzy sets using

Table 5: Pairwise comparison matrix

	Factor1	Factor2	Factor3	Factor4
Factor1	1	5	4	7
Factor2	1/5	1	1/2	3
Factor3	1/4	2	1	3
Factor4	1/7	1/3	1/3	1

Table 6: Fuzzy AHP scale

Intensity of Importance	Definition	Fuzzy sets
1	Equal	(1,1,1)
3	Moderate	(2,3,4)
5	Strong	(4,5,6)
7	Very strong	(6,7,8)
9	Extreme importance	(9,9,9)
2	Intermediate values	(1,2,3)
4	Intermediate values	(3,4,5)
6	Intermediate values	(5,6,7)
8	Intermediate values	(7,8,9)

the below formula-

$$a_{ij} = (l_{ij}, m_{ij}, u_{ij})$$

$$a_{ij}^{-1} = (\frac{1}{l_{ij}}, \frac{1}{m_{ij}}, \frac{1}{u_{ij}})$$

where l,j and u are the first, second, and third components of the fuzzy set.

- In our implementation, we used a a dictionary to create comparison matrix

$$dict1 = 1 : [1, 1, 1], 2 : [1, 2, 3], 3 : [2, 3, 4], 4 : [3, 4, 5],$$

$$5 : [4, 5, 6], 6 : [5, 6, 7], 7 : [6, 7, 8], 8 : [7, 8, 9], 9 : [9, 9, 9]$$

- The output obtained after using the dictionary, dict1 is described in Table 7.

Table 7: Pairwise comparison matrix

	Factor1	Factor2	Factor3	Factor4
Factor1	(1,1,1)	(4,5,6)	(3,4,5)	(6,7,8)
Factor2	(1/6,1/5,1/4)	(1,1,1)	(1/3,1/2,1/1)	(2,3,4)
Factor3	(1/5,1/4,1/3)	(1,2,3)	(1,1,1)	(2,3,4)
Factor4	(1/8,1/7,1/6)	(1/4,1/3,1/2)	(1/4,1/3,1/2)	(1,1,1)

4.3.1.2 Fuzzy weights calculation:

The following steps are performed to find the fuzzy weights at each level of hierarchy.

Algorithm 1 Fuzzy weights calculator

```
procedure FUZZYWEIGHT(matrix)                  ▷ function to find fuzzy weights
    number = len(matrix)                       ▷ number of criteria
    terms = 3
    sumArray = numpy.zeros(3)                  ▷ Create a sumArray
    inverseArray = numpy.zeros(3)              ▷ Create a inverseArray
    gm = numpy.ones((number, 3))               ▷ create a new 2D array
    for i in range(number): do
        for j in range(terms): do
            for k in range(number): do
                gm[i][j]∗ = matrix[i][k][j]
                gm[i][j] = (gm[i][j])^1/n      ▷ raising to power of 1/n
    for i in range(terms): do                  ▷ Find sum
        for j in range(n): do
            sumArray[i]+ = gm[j][i]
    inverseArray = 1/ sumArray
    Multiply the gm array and inverserArray
    Defuzzification to get Crisp numerical values.
```

- Step 1: Find the Geometric mean $\tilde{r_i}$ of fuzzy comparison values using the below formula:

$$\tilde{r_i} = (\prod_{i=1}^{n} (\tilde{a_{ij}})^{1/n}$$

Table 8 describes the calculated fuzzy geometric mean for all rows.

– Fuzzy Geometric Mean for row 1 in table 5 :

$$\tilde{r_1} = (1 * 4 * 3 * 6)^{\frac{1}{4}}, (1 * 5 * 4 * 7)^{\frac{1}{4}}, (1 * 6 * 5 * 8)^{\frac{1}{4}}$$

– Fuzzy Geometric Mean for row 2 in table 5:

$$\tilde{r_2} = (\frac{1}{6} * 1 * \frac{1}{3} * 2)^{\frac{1}{4}}, (\frac{1}{5} * 1 * \frac{1}{2} * 3)^{\frac{1}{4}}, (\frac{1}{4} * 1 * 1 * 4)^{\frac{1}{4}}$$

– Fuzzy Geometric Mean for row 3 in table 5:

$$\tilde{r_3} = (\frac{1}{5} * 1 * 1 * 2)^{\frac{1}{4}}, (\frac{1}{4} * 2 * 1 * 3)^{\frac{1}{4}}, (\frac{1}{3} * 3 * 1 * 4)^{\frac{1}{4}}$$

– Fuzzy Geometric Mean for row 4 in table 5:

$$\tilde{r_4} = (\frac{1}{8} * \frac{1}{4} * \frac{1}{4} * 1)^{\frac{1}{4}}, (\frac{1}{7} * \frac{1}{3} * \frac{1}{3} * 1)^{\frac{1}{4}}, (\frac{1}{6} * \frac{1}{2} * \frac{1}{2} * 1)^{\frac{1}{4}}$$

Table 8: fuzzy geometric mean

	Fuzzy geometric mean $\tilde{r_i}$
Factor1	(2.91, 3.44, 3.94)
Factor2	(0.58,0.74,1)
Factor3	(0.80,1.11, 1.41)
Factor4	(0.30,0.35,0.45)

- Step 2: For each $\tilde{r_i}$, find the direct sum using the below formula.

$$\begin{aligned}\sum_{r_i}^{n} &= r_{i1} \oplus r_{i2} \oplus ...... \oplus r_{in} \\ &= (4.58, 5.64, 6.80)\end{aligned}$$

$$\tilde{A}_1 \oplus \tilde{A}_2 = (l_1, m_1, u_1) \oplus (l_2, m_2, u_2) = ((l_1 + l_2, m_1 + m_2, u_1 + u_2)$$

- Step 3: For each $\tilde{r_i}$, find its inverse:

$$\sum_{r_i}^{n} = (r_{i1} \oplus r_{i2} \oplus ...... \oplus r_{in})^{-1}$$

In the above equation:

$$\tilde{w_i} = l\tilde{w_i}, m\tilde{w_i}, u\tilde{w_i}$$
$$l\tilde{w_i} = \text{least possible value}$$
$$u\tilde{w_i} = \text{maximum possible value}$$

$$(l, m, u)^{-1} = (\frac{1}{u}, \frac{1}{m}, \frac{1}{l})$$

- Step 4: Calculate relative fuzzy weights as depicted in table 9 using below formula:
$$\tilde{w_i} = \tilde{r_i} \otimes (\tilde{r_1} \oplus \tilde{r_2} \oplus \tilde{r_3} \oplus \tilde{r_n})^{-1}]$$

Table 9: Relative fuzzy weight table

	Relative fuzzy weight $\tilde{w_i}$	
Factor1	$(2.91, 3.44, 3.94) \otimes (\frac{1}{6.80}, \frac{1}{5.64}, \frac{1}{4.58})$	$(0.428, 0.610, 0.859)$
Factor2	$(0.58, 0.74, 1) \otimes (\frac{1}{6.80}, \frac{1}{5.64}, \frac{1}{4.58})$	$(0.085, 0.131, 0.218)$
Factor3	$(0.80, 1.11, 1.41) \otimes (\frac{1}{6.80}, \frac{1}{5.64}, \frac{1}{4.58})$	$(0.117, 0.196, 0.309)$
Factor4	$(0.30, 0.35, 0.45) \otimes (\frac{1}{6.80}, \frac{1}{5.64}, \frac{1}{4.58})$	$(0.044, 0.063, 0.099)$

- Step 5: Defuzzification: Use the below formula to get crisp numerical values M_i for each criteria as described in table 10:

$$M_i = \frac{(l\tilde{w_i}) \oplus (m\tilde{w_i}) \oplus (u\tilde{w_i})}{3}$$

$$\text{Center of Area} = \frac{l + m + u}{3}$$

- Step 6: Find total as depicted in Table 11:

$$Total = 0.633 + 0.145 + 0.207 + 0.068$$
$$if(total) > 1 \Rightarrow \text{Normalize } M_i \text{ using below formula}$$

$$f_i = \frac{M_i}{\sum_{i=1}^{n} M_i}$$

Table 10: Defuzzification

	$\widetilde{M}_i$
Factor1	0.633
Factor2	0.145
Factor3	0.207
Factor4	0.068

Table 11: Normalization

	Normalized weight
Factor1	$\frac{0.633}{1.058} = 0.601$
Factor2	$\frac{0.145}{1.058} = 0.138$
Factor3	$\frac{0.207}{1.058} = 0.197$
Factor4	$\frac{0.068}{1.058} = 0.065$

4.3.2 Calculation of global weighs for each criteria

In wildfire prediction, we need to consider multiple sub-criteria of various criteria. For example, weather would have multiple sub-criteria like pressure, humidity, temperature etc. The analytical hierarchy process is repeated for each criteria to find the local weights of its sub-criteria. The local weights of sub-criteria are then multiplied with global weights of criteria to find the global weights. The steps that should be followed are described in figure 16. These values are used to find the final ranking of various alternatives.

4.3.3 Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS)

We used TOPSIS that is a multi-criteria decision-making method, to rank the regions according to fire risks. The steps performed for decision making using TOPSIS

Criteria	Sub -Criteria	Weight of Criteria	Local Weights of Sub-Criteria	Global Weights of Criteria
Criteria 1	Sub -Criteria 1.1 Sub -Criteria 1.1	W1	W11 W12	W1 * W11 W1 * W12
Criteria 2	Sub -Criteria 2.1 Sub -Criteria 2.2	W2	W21 W22	W2 * W21 W2 * W22
Criteria 3	Sub -Criteria 3.1 Sub -Criteria 3.2	W3	W31 W32	W3 * W31 W3 * W32
Criteria 4	Sub -Criteria 4.1 Sub -Criteria 4.2	W4	W41 W42	W4* W41 W4 * W42

Figure 16: Calculation of global weights of criteria

are described below:

- Step 1: Create a M * N Matrix: Create a M * N matrix where M denotes the number of criteria and N denotes the number of alternatives.

$$(a_{ij})_{M*N}$$

- Step 2: Resolve linguistic factors Before applying the weights, the linguistics terms should be converted into a scale that can be compared. For example, if there is a column depicting weather condition as cloudy or clear, assign the following values:

$$Cloudy : 5$$
$$Clear : 2$$

- Step 3: Normalize the matrix using the below formula:

$$(a_{ij}) = \frac{a_{ij}}{\sqrt{\sum_{i=1}^{M}(a_{ij})^2}}$$

- Step 4: Calculate the weighted normalized matrix using weights obtained by the previous section.

$$\chi_{ij} = \omega_{ij} * \omega_j$$

- Step 5: For each column, find the maximum and minimum:

$$\chi_j^b = max(\chi_{ij})$$
$$\chi_j^w = min(\chi_{ij}))$$

- Step 6: Categorize the criteria into cost and benefit criteria.

  Cost: The values of these criteria should be less. The greater the criteria value, the more its preference. For example, Humidity

$$\chi_j^b = min(\chi_{ij})$$
$$\chi_j^w = max(\chi_{ij}))$$

  Benefit Criteria: The value of these criteria should be higher. The greater the criteria value, the less its preference. For example, Temperature

$$\chi_j^b = max(\chi_{ij})$$
$$\chi_j^w = min(\chi_{ij}))$$

- Step 7: Find the Euclidean distance between the best/worst alternative and the target alternative. The Euclidean distance as described in figure 17 is the shortest distance between two points in space.

$$\text{positiveIdeal} = \sum_{j=1}^{N} (\chi_{ij-\chi_j^b})^2$$
$$\text{negativeIdeal} = \sum_{j=1}^{N} (\chi_{ij-\chi_j^w})^2$$

- Step 8: Find the TOPSIS score

$$\text{TOPSIS score} = \frac{negativeIdeal}{positiveIdeal + negativeIdeal}$$

- Step 9: Rank the alternatives using the TOPSIS score.

$$\text{Ranking} = TOPSISscore.sortValues()$$

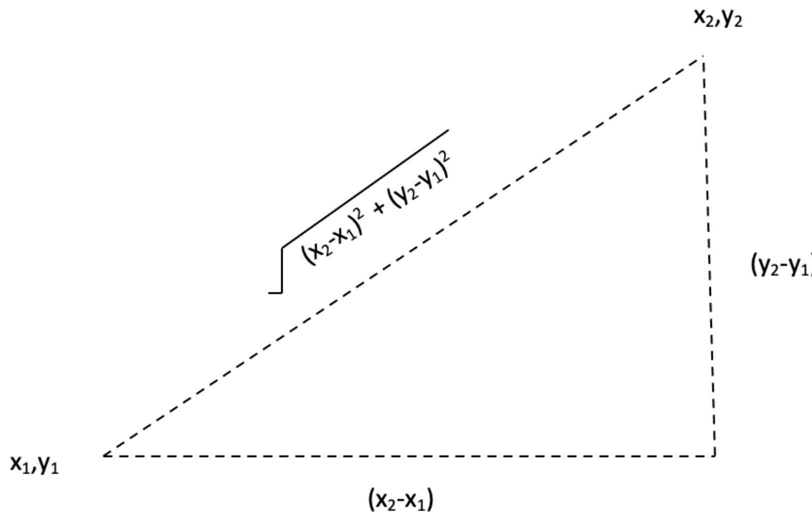


Figure 17: Euclidean distance

4.4 Integration with smart city's resiliency plan

We will integrate the fire risk into the city's resiliency plan so that city can prepare for a high-risk zone. The city administrates for the regions that have a higher rank in future wildfire ranking would get an email notification. The city can improve the infrastructure according to the risk zones using proper design and construction. We used Simple Mail Transfer Protocol (SMTP) email server to send the email.

$$server = SMTP('smtp.gmail.com', 587)$$

$$server.starttls()$$

CHAPTER 5

Experimental Results

This section discusses the experiments that are performed on the California cities data. We derived the data set about California cities from multiple sources as described in the table.

5.1 California cities

Experiments are performed on the 459 California cities. The topography, vegetation, and location details are obtained after performing the analysis on shapefiles. Figure 18 represents the California boundary map, and the location of various cities is depicted in color dots.

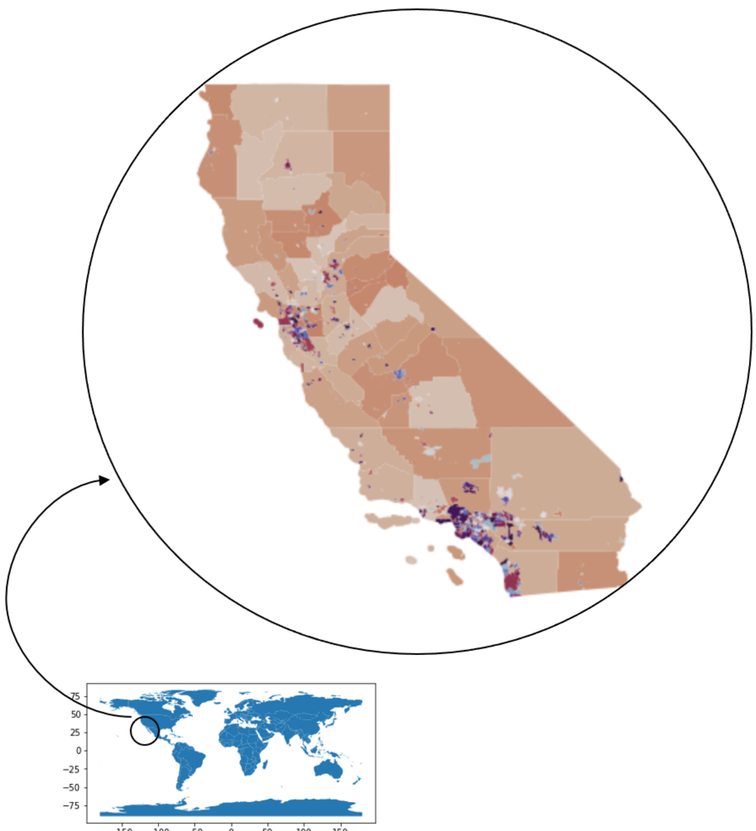


Figure 18: California boundaries map

The CSV file obtained from simplemap.com contains information about all the cities of California. Topography data is extracted for all cities of California and stored in a data frame. The shapefile data frame is merged with the CSV file data frame. The various attributes of the merged data frame are depicted in the figure 19.

	city	city_ascii	state_id	state_name	county_fips	county_name	lat	lng	population	density	source	military	incorporated	timezone	ranking	zips	id	OBJECTID	COUNTY
0	Los Angeles	Los Angeles	CA	California	6037	Los Angeles	34.1139	-118.4068	12750807	3276	polygon	False	True	America/Los_Angeles	1	90291 90293 90292 91316 91311 90037 90031 9000...	1840020491	240.0	Los Angeles County
1	San Francisco	San Francisco	CA	California	6075	San Francisco	37.7562	-122.4430	3592294	7256	polygon	False	True	America/Los_Angeles	1	94130 94131 94132 94133 94134 94109 94108 9410...	1840021543	370.0	San Francisco County
2	San Diego	San Diego	CA	California	6073	San Diego	32.8312	-117.1225	3220118	1686	polygon	False	True	America/Los_Angeles	1	92109 92108 92103 92111 92154 92110 92115 9214...	1840021990	367.0	San Diego County
3	Riverside	Riverside	CA	California	6065	Riverside	33.9381	-117.3948	2107852	1574	polygon	False	True	America/Los_Angeles	1	92508 92503 92501 92505 92504 92507 92506 9250...	1840020551	352.0	Riverside County

Figure 19: Merged dataframe

5.1.1 Weather data

The weather of latitude and longitude is forecasted using the API call to openweathermap.org. OpenWeatherMap is an online service to obtain the world's weather data using Application program interface (API) calls. Using the following steps, weather of California cities is forecasted for next fifteen days:

- Step 1: Registration:

  We registered on openweathmap.org and subscribed to the "FREE" version that allows us to make 1,000 API calls/day. After successful registration, a unique API key was obtained. The format of the API call is :

$$api.openweathermap.org/data/2.5/forecast?$$
$$+$$
$$lat = \{lat\}\&lon = \{lon\}\&appid = \{APIkey\}$$

- Step 2: Latitude and Longitude:

We retrieved the latitude and longitude of each city from the data frame obtained in the previous section. We made the 459 API call to obtain the weather forecast for 459 California cities.

- Step 3: JavaScript Object Notation (JSON):

  Figure 20 depicts the output of the openweathmap API calls. We observed that the data is obtained in JSON format.


```
{"cod":"200","message":0,"cnt":40,"list":[{"dt":1620010800,"main":
{"temp":18.17,"feels_like":16.96,"temp_min":18.17,"temp_max":18.27,"pressure":1002,"sea_level":1002,"grnd_level":909,"humidity":35,"temp_kf":-0.1},"weather":
[{"id":800,"main":"Clear","description":"clear sky","icon":"01n"}],"clouds":{"all":0},"wind":
{"speed":8.31,"deg":263,"gust":13.41},"visibility":10000,"pop":0,"sys":{"pod":"n"},"dt_txt":"2021-05-03 03:00:00"},{"dt":1620021600,"main":
{"temp":17.55,"feels_like":16.22,"temp_min":16.32,"temp_max":17.55,"pressure":1004,"sea_level":1004,"grnd_level":910,"humidity":33,"temp_kf":1.23},"weather":
[{"id":800,"main":"Clear","description":"clear sky","icon":"01n"}],"clouds":{"all":0},"wind":
{"speed":5.94,"deg":251,"gust":9.89},"visibility":10000,"pop":0,"sys":{"pod":"n"},"dt_txt":"2021-05-03 06:00:00"},{"dt":1620032400,"main":
{"temp":16.08,"feels_like":14.53,"temp_min":15.03,"temp_max":16.08,"pressure":1007,"sea_level":1007,"grnd_level":910,"humidity":30,"temp_kf":1.05},"weather":
[{"id":800,"main":"Clear","description":"clear sky","icon":"01n"}],"clouds":{"all":1},"wind":
{"speed":3.96,"deg":265,"gust":6.16},"visibility":10000,"pop":0,"sys":{"pod":"n"},"dt_txt":"2021-05-03 09:00:00"},{"dt":1620043200,"main":
{"temp":13.4,"feels_like":11.74,"temp_min":13.4,"temp_max":13.4,"pressure":1010,"sea_level":1010,"grnd_level":911,"humidity":36,"temp_kf":0},"weather":
[{"id":800,"main":"Clear","description":"clear sky","icon":"01n"}],"clouds":{"all":3},"wind":{"speed":2.04,"deg":284,"gust":2.6},"visibility":10000,"pop":0,"sys":
{"pod":"n"},"dt_txt":"2021-05-03 12:00:00"},{"dt":1620054000,"main":
{"temp":17.66,"feels_like":16.13,"temp_min":17.66,"temp_max":17.66,"pressure":1011,"sea_level":1011,"grnd_level":914,"humidity":25,"temp_kf":0},"weather":
[{"id":800,"main":"Clear","description":"clear sky","icon":"01d"}],"clouds":{"all":9},"wind":{"speed":1.55,"deg":5,"gust":3.1},"visibility":10000,"pop":0,"sys":
{"pod":"d"},"dt_txt":"2021-05-03 15:00:00"},{"dt":1620064800,"main":
```


Figure 20: JSON weather data

- Step 4: Processing the JSON data:

  The JSON data obtained in the previous step contains various attributes like time, city, latitude, longitude, country, population, temperature, pressure, humidity, sky, sky description, clouds, wind speed, wind direction, etc. We processed the JSON data to retrieve only the relevant columns as described in figure 21 in and stored them in multiple lists. The data-frame for weather data is represented in figure 22.

- Step 5: Filtering weather for a specific time: The weather API predicts the weather at three-hour intervals at 9:00, 12:00, 15:00, 18:00, 21:00 every day. The hottest time of day is not noon as the atmosphere's temperature is not increased by the incoming sun rays. Earth is responsible for heating of atmosphere as it emits heat radiations. The re-radiation starts later in the day, around 15:00. So, we will use the weather information received at 3:00 PM to predict the chances of wildfires.

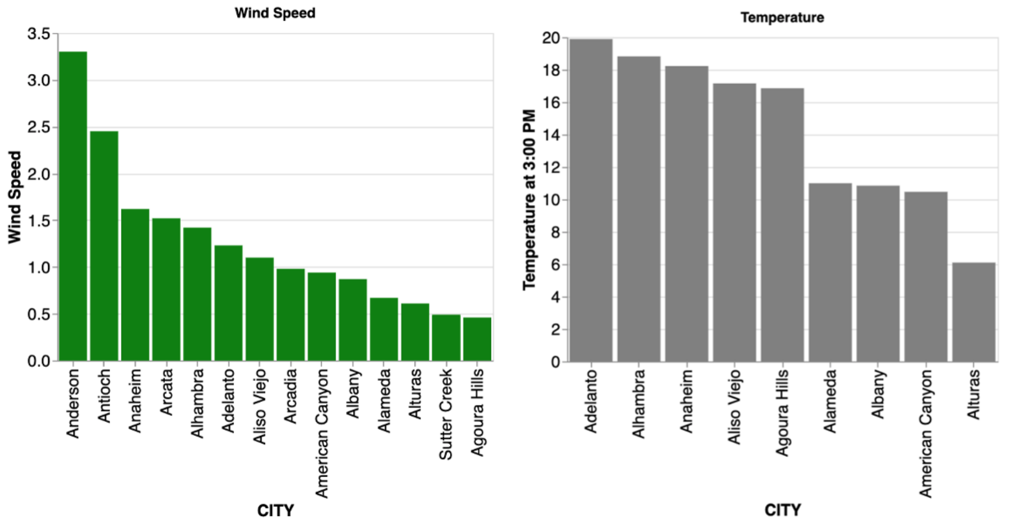


Figure 21: Temperature and pressure from weather forecast

	city	Time	population	temperature	pressure	humidity	wind_speed	wind_direction
35	Adelanto	2021-05-07 15:00:00	31765	19.89	1012	23	3.32	221
35	Agoura Hills	2021-05-07 15:00:00	20330	16.86	1015	62	0.99	114
35	Alameda	2021-05-07 15:00:00	73812	11.00	1019	77	1.38	237
35	Albany	2021-05-07 15:00:00	18539	10.85	1019	71	1.82	274
35	Alhambra	2021-05-07 15:00:00	83089	18.82	1015	55	1.48	193
35	Aliso Viejo	2021-05-07 15:00:00	47823	17.16	1015	68	3.06	158
35	Alturas	2021-05-07 15:00:00	2827	6.10	1019	64	3.23	316
35	Sutter Creek	2021-05-07 15:00:00	2501	12.43	1017	53	1.27	268
35	American Canyon	2021-05-07 15:00:00	19454	10.47	1019	66	1.69	267
35	Anaheim	2021-05-07 15:00:00	336265	18.23	1015	62	1.75	183

Figure 22: Processed API weather data

5.1.2 Vegetation data

Vegetation data for each city in California is obtained by performing data analysis on various kinds of vegetation data described in chapter 4. The vegetation data for the relevant cities are obtained and merged with the weather data of the corresponding city.

5.1.3 Population density

Population of the city i.e the number of inhabitants of a city can be obtained by the openweathmap API call output or from the data set of the United States Census Bureau. Figure 23 is a plot of the California boundaries combined with Census department's county wise population data. Population density of each region is calculated by dividing its population by the total area of that region.

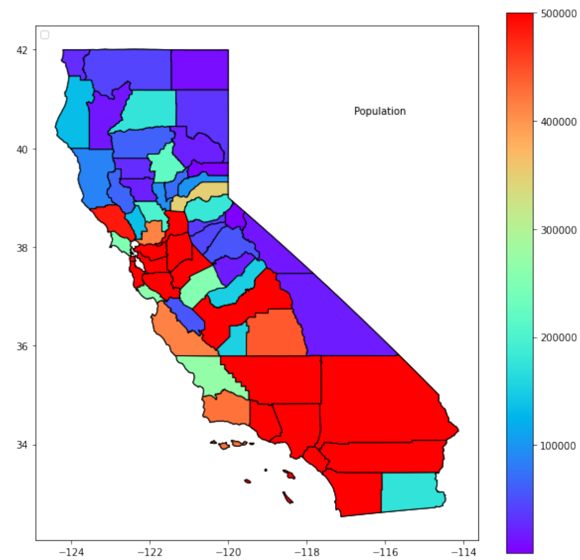


Figure 23: California county population map

5.1.4 Combining the data

All the data frames are combined together using city names as the primary key. A snapshot of the combine data frame for ten cities of California is represented in Figure 22.

	city	temperature	pressure	humidity	wind_speed	Alfalfa & Hay_percentage	Deciduous Forest_percentage	Evergreen Forest_percentage	Barren_percentage	Grassland_percentage	Wetlands_percentage	max_elevation	density
0	Adelanto	19.89	1012	23	3.32	0.022024	0.000632	1.814040	10.975012	0.871968	0.040986	3493	248
1	Agoura Hills	16.86	1015	62	0.99	0.066988	0.002494	4.843746	3.594006	9.303955	0.087450	3032	1000
2	Alameda	11.00	1019	77	1.38	0.300074	0.001452	2.132827	0.052947	41.812059	1.223712	1242	2868
3	Albany	10.85	1019	71	1.82	0.300074	0.001452	2.132827	0.052947	41.812059	1.223712	1242	4250
4	Alhambra	18.82	1015	55	1.48	0.066988	0.002494	4.843746	3.594006	9.303955	0.087450	3032	4237
5	Aliso Viejo	17.16	1015	68	3.06	0.001428	NaN	1.009126	0.316972	8.033657	0.652157	1732	2839
6	Alturas	6.10	1019	64	3.23	3.263169	0.000025	27.245897	4.569542	20.250476	1.275018	3004	350
7	American Canyon	10.47	1019	66	1.69	0.009883	1.105416	13.648016	0.017469	17.042326	1.159820	84	1298
8	Anaheim	18.23	1015	62	1.75	0.001428	NaN	1.009126	0.316972	8.033657	0.652157	1732	2688

Figure 24: Snapshot of data framework

5.1.5 Matrix formation

The data is converted into a M * N matrix where M denotes the number of criteria and N denotes the number of alternatives. The matrix (a_{ij} is input to our model.

$$(a_{ij})_{M*N}$$

5.2 AHP fuzzy weight calculation

The model asks the decision-maker to specify the degree of importance of one criteria to all other criteria. The decision-maker can use the data analysis performed in the previous chapter to determine each criterion's importance. The assigned weights are converted into fuzzy weights and act as input to our model. Figure 25 describes the process used to convert local weights of each criteria into into global weights.

Criteria	Weight	Sub-criteria	Local Weight	Global Weight
Weather	W	Temperature	W_1	$W * W_1$
		Pressure	W_2	$W * W_2$
		Humidity	W_3	$W * W_3$
		Wind Speed	W_4	$W * W_4$
Vegetation	V	Alfalfa & Hay %	V_1	$V * V_1$
		Deciduous Forest %	V_2	$V * V_2$
		Evergreen %	V_3	$V * V_3$
		Barren %	V_4	$V * V_4$
		Grassland %	V_5	$V * V_5$
		Wetland %	V_6	$V * V_6$
Geography	G	-		G
Population Density	P	-		P

Figure 25: Global weights calculation

5.3 Weighted normalized matrix

The global weights are used to find the find the weighed normalized matrix as shown in Figure 26.

5.4 California cities wildfire risk ranking

Our model ranks the California cities according to wildfire risks as depicted in figure 27. The highest fire risk cities can use this information to prepare for the resources required to prevent and control the wildfires.

```
The weighted normalized
                    0         1         2   ...         9        10        11
Adelanto      0.021892  0.003092  0.000756  ...  0.000007  0.006900  0.000168
Agoura Hills  0.018678  0.003101  0.001938  ...  0.000014  0.005989  0.000678
Alameda       0.012107  0.003113  0.002530  ...  0.000199  0.002453  0.001945
Albany        0.011942  0.003113  0.002333  ...  0.000199  0.002453  0.002882
Alhambra      0.020715  0.003101  0.001807  ...  0.000014  0.005989  0.002873
...                ...       ...       ...  ...       ...       ...       ...
Woodland      0.013208  0.003110  0.002168  ...  0.000211  0.000683  0.001035
Yorba Linda   0.020296  0.003101  0.001873  ...  0.000106  0.003421  0.000892
Yreka         0.007022  0.003126  0.002234  ...  0.000049  0.008482  0.000197
Yuba City     0.014078  0.003110  0.001708  ...  0.000653  0.003263  0.001177
Yucaipa       0.019570  0.003101  0.001018  ...  0.000007  0.006900  0.000498
```

Figure 26: Weighted normalized matrix of California cities

City	Rank
Avenal	1.0
Palm Desert	2.0
Indian Wells	3.0
Hanford	4.0
Ridgecrest	5.0
Taft	6.0
Bakersfield	7.0
Wasco	8.0
Fowler	9.0
El Cajon	10.0
Santee	11.0

Figure 27: California cities wildfire risk ranking for smaller dataset

5.5 Model Accuracy

The model is checked against the historical data and current data. The OpenWeatherMap does not return historical weather record in "Free subscription" API. So, we fetched the weather for particular data using www.wunderground.com. Some information of weather is also obtained from The National Oceanic and Atmospheric Administration.

5.5.1 Historical data

Experiments are performed on the wildfire incident data for the year 2014 for 230 regions, figure 28 in four counties, Los Angeles, San Diego, Shasta County, and Mariposa county as plotted in figure 29.

```
               city  ...          id
0       Los Angeles  ...  1840020491
11       Long Beach  ...  1840020490
19        Lancaster  ...  1840020516
22    Santa Clarita  ...  1840021864
33         Glendale  ...  1840020483
...             ...  ...         ...
1315         Wawona  ...  1840024756
1392      Fish Camp  ...  1840024753
1416       Hornitos  ...  1840026811
1418   Coulterville  ...  1840028088
1464   Buck Meadows  ...  1840022542

[230 rows x 17 columns]
```

Figure 28: California cities from four counties

From the wildfire history, we extracted following information for Colby fire:

- Name of the fire: Colby fire
- Fire data: January 16, 2014
- Vegetation: Details are retrieved from our data set
- Topography: Details are retrieved from our data set
- Weather: details for all the cities are considered by taking average temperature, average humidity, and average pressure for the city in January 2014.

The model ranked 27 cities near Angeles national forest at top 40 places and the cities in Shasta County at the bottom of the ranking. The cities are matched with Colby fire locations as depicted in figure. figure 30.

5.5.2 Current data

The current fire risk ranking for all California cities is described in figure 27. The data is collected on May 1, 2021. Most of the cities in the ranking lie in the periphery or near the boundary of current wildfires. The source of current wildfires name and

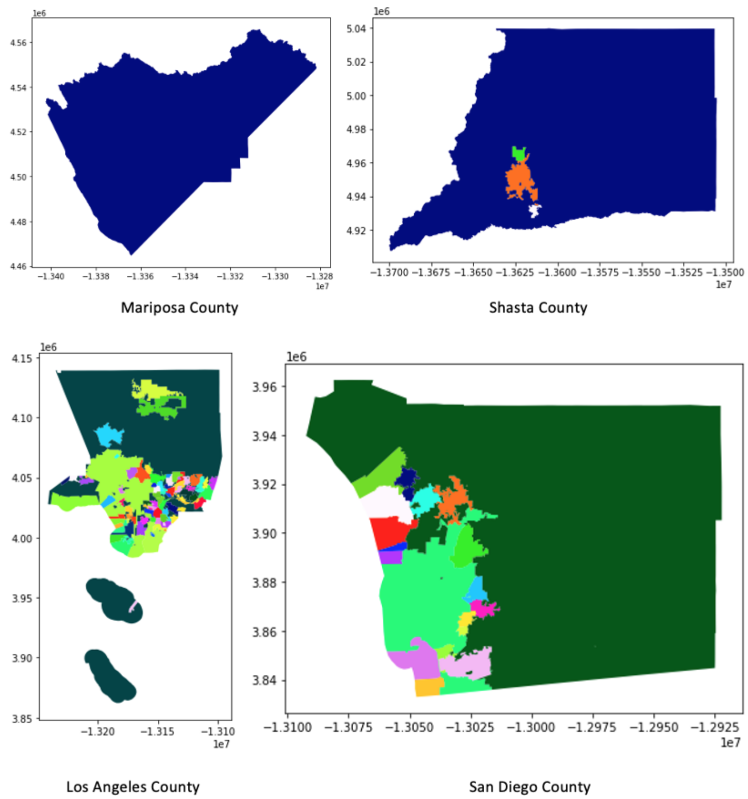


Figure 29: County map

Figure 30: Colby fire 2014 location, source:Wikipedia

description is Wikipedia and details of the current wildfires are explained in figure 31.

Another experiment performed on weather conditions for May 14, 2021, is de-

Name	County	Acres	Start date	Containment date	Notes	Ref [hide]
Owens	Kern	1,512	May 1	85% contained as of May 5		[12]
Southern	San Diego	5,366	May 1	90% contained as of May 5		[13]

Figure 31: Current wildfire season in California, source:Wikipedia

scribed in the figure 32.

City	Rank
Fresno	1.0
Jackson	2.0
Banning	3.0
Weed	4.0
Redding	5.0
Anderson	6.0
Arcadia	7.0
Bishop	8.0
Pomona	9.0
Bell	10.0
Morgan Hill	11.0
Saratoga	12.0
Beverly Hills	13.0
Walnut	14.0
Atascadero	15.0
El Cajon	16.0
Agoura Hills	17.0
Santee	18.0
Los Alamitos	19.0
La Habra Heights	20.0

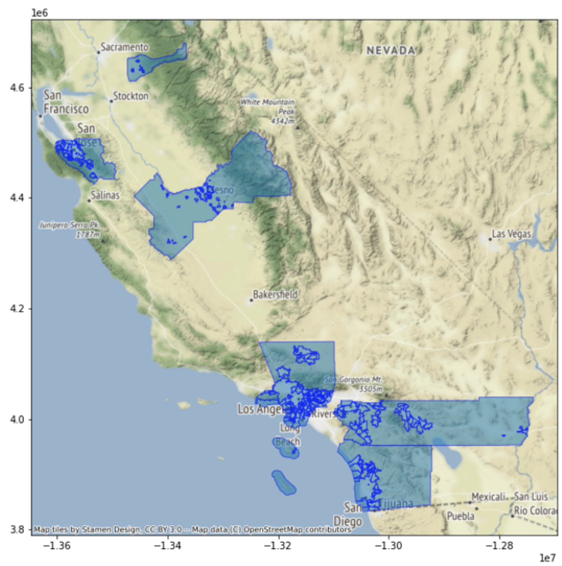


Figure 32: Wildfire risk in California

The cites that appear in the top 20 results match the cities in the fire hazard severity zone map of the California fire department of forestry and fire protection. The cities that appeared in top fire risk zones are located in Amador, Fresno, San Diego, Los Angeles, Riverside, and Santa Clara counties. Figure 33 represents the fire hazard severity zone map of the California fire department of forestry and fire protection.

5.6 Integration with Smart City

We know that the resources required to contain the wildfires are not unlimited. So, this model aims to warn city authorities about future fire risks. Fire warnings are sent to cities that have high chances of wildfire in the coming few days. Fire

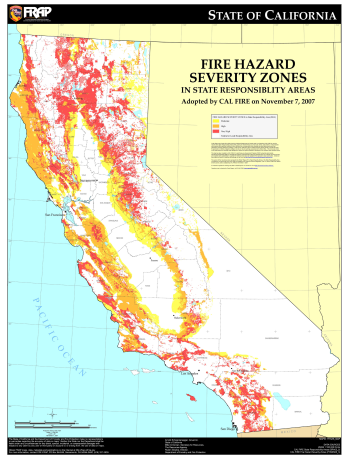
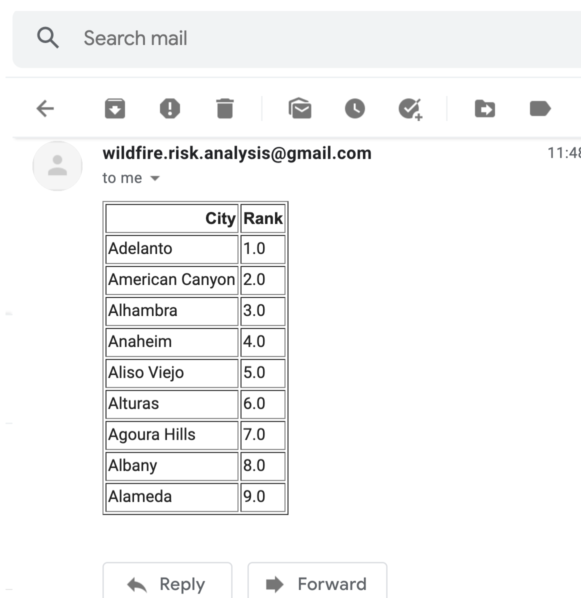


Figure 33: Fire hazard severity zone map, Source: CA.gov

warnings are sent using SMTP email server to the registered emails. Email received using our alert system is represented in figure 34. The civil authorities can issue a warning through Emergency Alert System (EAS) to inform the city dwellers about fire risk in their area.


Search mail

wildfire.risk.analysis@gmail.com

to me

11:4

City	Rank
Adelanto	1.0
American Canyon	2.0
Alhambra	3.0
Anaheim	4.0
Aliso Viejo	5.0
Alturas	6.0
Agoura Hills	7.0
Albany	8.0
Alameda	9.0

Reply Forward

Figure 34: Alert email snapshot

CHAPTER 6

Conclusion and Future Work

In this research, we quantified fire risk in California by using a multi-criteria decision-making technique and ranked the cities of California based on their risk of wildfires. We incorporated fuzzy set theory to handle the imprecision and uncertainty. This model can be used for wildfire risk analysis in different states. It can also be helpful for developing countries where resources are limited to predict wildfire risks and for preparedness planning. This fire risk model can be integrated with the housing models of a smart city to determine whether you live in a wildfire zone or not. Knowing the fire risks, a city can use this model as a tactical guide to design the houses in a fire risk zone and build multiple fire stations in critical zones. City authorities can incorporate information technology to send emergency fire alerts.

However, wildfires are dynamic in nature and predicting the risk of wildfires with 100% accuracy is not feasible. As a part of future work, we can improve the model by incorporating additional causative factors of wildfires into our model. We can add sensor data and satellite data to find the current atmospheric condition of a region. The population factor can be improved by incorporating the density of electricity power lines in that area. Various fuel moisture codes like fine fuel moisture code can be incorporated to provide numeric fire intensity ratings. The model can also be integrated with already developed fire prediction model for risk analysis in different regions.

LIST OF REFERENCES

[1] A. Alonso-Betanzos, O. Fontenla-Romero, B. Guijarro-Berdiñas, E. Hernández-Pereira, J. Canda, E. Jiménez, J. L. Legido Soto, S. Muñiz, C. Paz-Andrade, and M. Paz-Andrade, ‘‘A neural network approach for forestal fire risk estimation.’’ in *Proceedings of the 15th European Conference on Artificial Intelligence*, 01 2002, pp. 643–647.

[2] R. Lagerquist, M. D. Flannigan, X. Wang, and G. A. Marshall, ‘‘Automated prediction of extreme fire weather from synoptic patterns in northern alberta, canada,’’ *Canadian Journal of Forest Research*, vol. 47, no. 9, pp. 1175–1183, 2017. [Online]. Available: https://doi.org/10.1139/cjfr-2017-0063

[3] A. Malik, M. R. Rao, N. Puppala, P. Koouri, V. A. K. Thota, Q. Liu, S. Chiao, and J. Gao, ‘‘Data-driven wildfire risk prediction in northern california,’’ *Atmosphere*, vol. 12, no. 1, 2021. [Online]. Available: https://www.mdpi.com/2073-4433/12/1/109

[4] Wikipedia contributors, ‘‘2020 california wildfires --- Wikipedia, the free encyclopedia,’’ https://en.wikipedia.org/w/index.php?title=2020_California_wildfires&oldid=994208447, 2020, [Online; accessed 15-December-2020].

[5] S. Kaplan and B. J. Garrick, ‘‘On The Quantitative Definition of Risk,’’ *Risk Analysis*, vol. 1, no. 1, pp. 1127, Mar. 1981. [Online]. Available: http://doi.wiley.com/10.1111/j.1539-6924.1981.tb01350.x

[6] A. Alonso-Betanzos, O. Fontenla-Romero, B. Guijarro-Berdiñas, E. Hernández-Pereira, M. Inmaculada Paz Andrade, E. Jiménez, J. Luis Legido Soto, and T. Carballas, ‘‘An intelligent system for forest fire risk prediction and fire fighting management in galicia,’’ *Expert Systems with Applications*, vol. 25, no. 4, pp. 545 –554, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417403000952

[7] H. Kumar, M. K. Singh, M. Gupta, and J. Madaan, ‘‘Moving towards smart cities: Solutions that lead to the smart city transformation framework,’’ *Technological Forecasting and Social Change*, vol. 153, p. 119281, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S004016251731394X

[8] Wikipedia contributors, ‘‘Fuzzy set --- Wikipedia, the free encyclopedia,’’ https://en.wikipedia.org/w/index.php?title=Fuzzy_set&oldid=993067551, 2020, [Online; accessed 15-December-2020].

[9] S. Zionts, "Mcdm—if not a roman numeral, then what?" *Interfaces*, vol. 9, no. 4, p. 94–101, Aug 1979. [Online]. Available: http://dx.doi.org/10.1287/inte.9.4.94

[10] Wikipedia contributors, "Decision-making --- Wikipedia, the free encyclopedia," 2020, [Online; accessed 15-December-2020]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Decision-making&oldid=994289678

[11] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "Knowledge discovery and data mining: Towards a unifying framework," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD'96. AAAI Press, 1996, p. 82–88.

[12] A. Camia, P. Barbosa, G. Amatulli, and J. San-Miguel-Ayanz, "Fire danger rating in the european forest fire information system (effis): Current developments," *Forest Ecology and Management*, vol. 234, p. S20, 11 2006.

[13] L. S. Bradshaw, *The 1978 national fire-danger rating system: technical documentation.* US Department of Agriculture, Forest Service, Intermountain Forest and Range . . . , 1984, vol. 169.

[14] C. Van Wagner, P. Forest, *et al.*, "Development and structure of the canadian forest fireweather index system," in *Can. For. Serv., Forestry Tech. Rep.* Citeseer, 1987.

[15] T. Cavazos, "Using Self-Organizing Maps to Investigate Extreme Climate Events: An Application to Wintertime Precipitation in the Balkans." *Journal of Climate*, vol. 13, no. 10, pp. 1718–1732, May 2000.

[16] M. Crimmins, "Synoptic climatology of extreme fire weather conditions across the southwest united states," *International Journal of Climatology*, vol. 26, pp. 1001 –1016, 06 2006.

[17] M. Pakdaman, S. S. Naghab, L. Khazanedari, S. Malbousi, and Y. Falamarzi, "Lightning prediction using an ensemble learning approach for northeast of iran," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 209, p. 105417, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364682620302236

[18] Wikipedia contributors, "Evapotranspiration --- Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Evapotranspiration&oldid=991872150, 2020, [Online; accessed 16-December-2020].

[19] Y. Safi and A. Bouroumi, "Prediction of forest fires using artificial neural networks," *Applied Mathematical Sciences*, vol. 7, pp. 271–286, 01 2013.

[20] H. Naganathan, S. Seshasayee, J. Kim, J.-S. Chou, and W. Chong, "Wildfire predictions: Determining reliable models using fused dataset," *Global Journal of Computer Science and Technology*, vol. 16, pp. 35–46, 11 2016.

[21] G. E. Sakr, I. H. Elhajj, G. Mitri, and U. C. Wejinya, "Artificial intelligence for forest fire prediction," in *2010 IEEE/ASME international conference on advanced intelligent mechatronics*. IEEE, 2010, pp. 13111316.

[22] G. Jakovljević, L. Gigovic, D. Sekulovic, and M. Regodić, "Gis multi-criteria analysis for identifying and mapping forest fire hazard: Nevesinje, bosnia and herzegovina," *Tehnicki Vjesnik*, vol. 25, 06 2018.

[23] A. Lapucci, S. Lombardo, M. Petri, and A. Santucci, "A kdd based multicriteria decision making model for fire risk evaluation," in *Conference on GIS-cience (AGILE)*. Citeseer, 2005.

[24] R. C. Rothermel, "How to predict the spread and intensity of forest and range fires," Jan 1983. [Online]. Available: https://www.fs.usda.gov/treesearch/pubs/24635