ne New Yor

NEW YORK, FRIDAY, SEPTEMBER 11, 2020

DISASTROU OF CLIMATE EVEN SLAMS CALIFORNIA

As Northwest Burns, Pleading for Help

Scientists Fear Fir Are Just the Start

Fast-moving wildfire in the Canadian Rockies' largest national park hits the town of Jasper

NEWS | EARTH

Human-sparked wildfires are more destructive than those caused by nature

They burn faster

While half of the July wildfires were caused by lightning, the other half were caused by people, Tucker said, warning people to respect the fire ban across Alberta's Forest Protection Area.

Nearly 1,100 Albertans forced from ho extreme heat fuels wildfires

Hot, dry c

pected to continue over the weekend

CBC News · Posted: Jul 19, 2024 5:38 PM MDT | Last Updated: J

dy Shows 84% of Wildfires Caused by Humans

the last 21 years, debris burning, arson and campfires have combined with climate change to the fire season much longer

Exploring Wildfire Distribution of Alberta: The Role of Human Activity and Fire Class Classification Models

Report by: Archana Senthil

Introduction

Wildfires have been rapidly increasing in frequency and severity around the world, with Alberta currently battling six active fires that have been burning since last year. While these fires are under control, their lasting impact on the ecosystem is undeniable. Wildfires are a natural part of Earth's ecological cycle and can be tolerated when caused by natural factors. However, wildfires sparked by human activities pose a much greater threat and disrupt delicate ecosystems.

To distinguish between **natural and human-induced wildfires**, a **detailed analysis** of wildfire data is necessary. **Historical wildfire records from Alberta (2006–2023), obtained from the Government of Alberta, provide key insights into fire trends, causes, and the impact of human causes.**

This dataset includes **26,552 recorded wildfire incidents** with **50 features**, covering:

- **Wildfire classification**
- **Start and end dates**
- **Environmental conditions** (wind speed, temperature, humidity, fuel type, etc.)
- **Geographical location and spread**
- **Cause of ignition**

By examining this extensive dataset, we can **identify trends, assess fire severity, and quantify the impact of human activities on wildfire occurrences.**

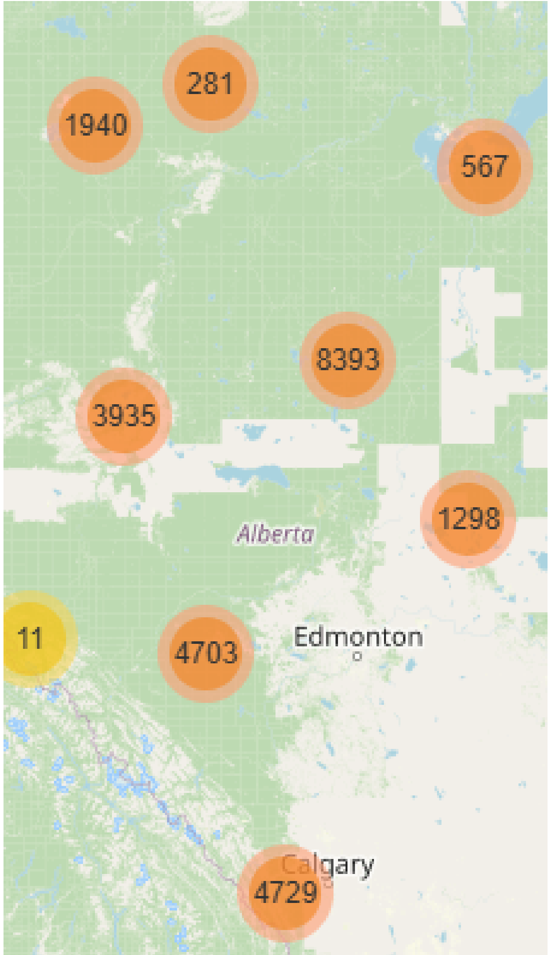
Data Preprocessing

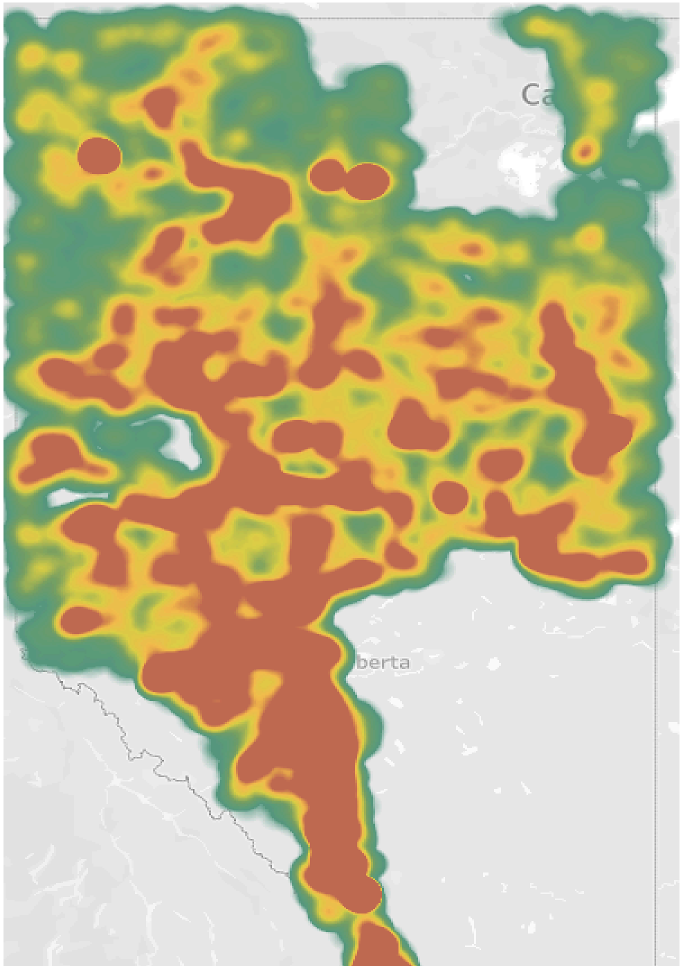
In the preprocessing stage, after reading the dataset in python, unused columns that wouldn't contribute to the analysis, such as various date-related columns, reagent type, and how the wildfire was notified, were removed. The dataset was then checked for null values, and rows with missing start dates were discarded as they lacked crucial information, including environmental conditions, which were necessary for the analysis. A new column was added to calculate the number of days the wildfire was active, which was determined by subtracting the start date from the end date. For the remaining missing values in environmental factors like temperature, humidity, wind speed, and fire spread rate, empty cells were filled with the average values for that month across the years, considering the seasonal variations in these factors. With these adjustments, the dataset is now clean, consistent, and ready for further analysis.

Objective 1: Understanding Wildfire Distribution

To achieve this objective, we will explore the key question: ***How are wildfires distributed across Alberta?***

Between 2006 and 2023, Alberta has experienced **26,552 wildfires**, impacting a perimeter over **6.4 million hectares** of land. These wildfires originate from both natural causes, such as lightning, and human-related activities, including industrial accidents, agricultural burns, and intentional fires.



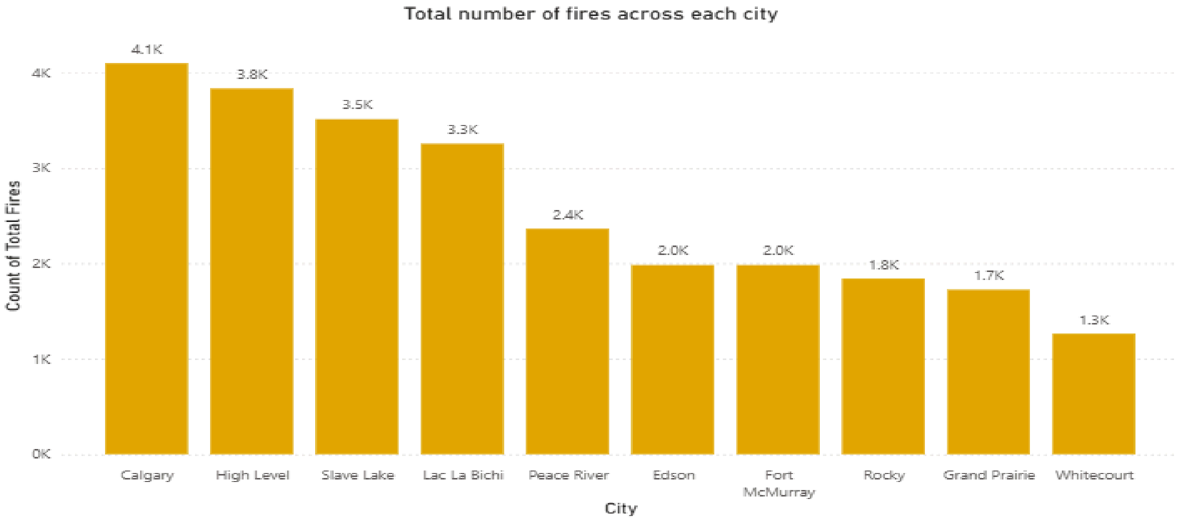


Understanding wildfire distribution is essential for developing effective mitigation strategies. A **geospatial analysis** of wildfire occurrences across Alberta highlights distinct patterns in wildfire frequency and location. The visualizations above illustrate the total number of wildfires in each sector, mapped using their geographic coordinates.

To create this geospatial representation, **Python's Folium library** was used, employing the **MarkerCluster** feature. This technique clusters wildfire occurrences based on latitude and longitude, allowing for better visualization as users zoom in and out of the map.The map on the right was developed in **Tableau,** showing the same visuals in the form of a heat map. The analysis indicates that the highest concentration of wildfires is found in **western and central Alberta**, likely due to **colder temperatures in the north** and **more open plains in the east and south**.

Further, wildfire distribution is classified according to the **forest area where ignition occurs**. This classification is determined by the **fire number**, where the first letter corresponds to a specific **forest region**. The affected regions include: **Calgary, Edson, High Level, Grand Prairie, Lac La Biche, Fort McMurray, Peace River, Rocky, Slave Lake, Whitehorn.**

To analyze the distribution of wildfires across these regions, a **column graph** was created using **Power BI**. The wildfire count for each forest area was extracted by comparing a dictionary of city names in Python with the fire number data. This visualization provides a clearer perspective on which **regions experience the highest frequency of wildfires**, aiding in targeted prevention and response efforts. The graph clearly shows that **Calgary** and **High Level** face the **most number of Wildfires** in Alberta, covering over **30%** of the total wildfires.



Objective 2: Assessing Human Impact on Wildfires

To evaluate the influence of human activities on wildfire occurrences, we will explore the following key questions: ***What are the primary causes of wildfires? To what extent do non-industrial human activities contribute to wildfire incidents?***

Wildfires can be broadly classified into two primary categories: **Natural Causes** – Primarily **lightning**, which is a natural part of forest ecosystems. **Human-Related Causes** – These include fires caused by: **Residential activities** (fires originating near homes in forested areas), **Recreational activities** (camping, berry picking, fishing, etc.), **Incendiary actions** (deliberate arson), **Government activities, Railroad incidents, Industrial accidents** (including oil & gas, power lines, and forestry), **Agricultural burns, Prescribed fires.**

While natural wildfires are **inevitable** and play a role in ecological balance, human-caused wildfires are **preventable**. Implementing stricter fire regulations and increasing public awareness can significantly reduce the number of human-induced fires.

Among human-related wildfires, three major contributors stand out:

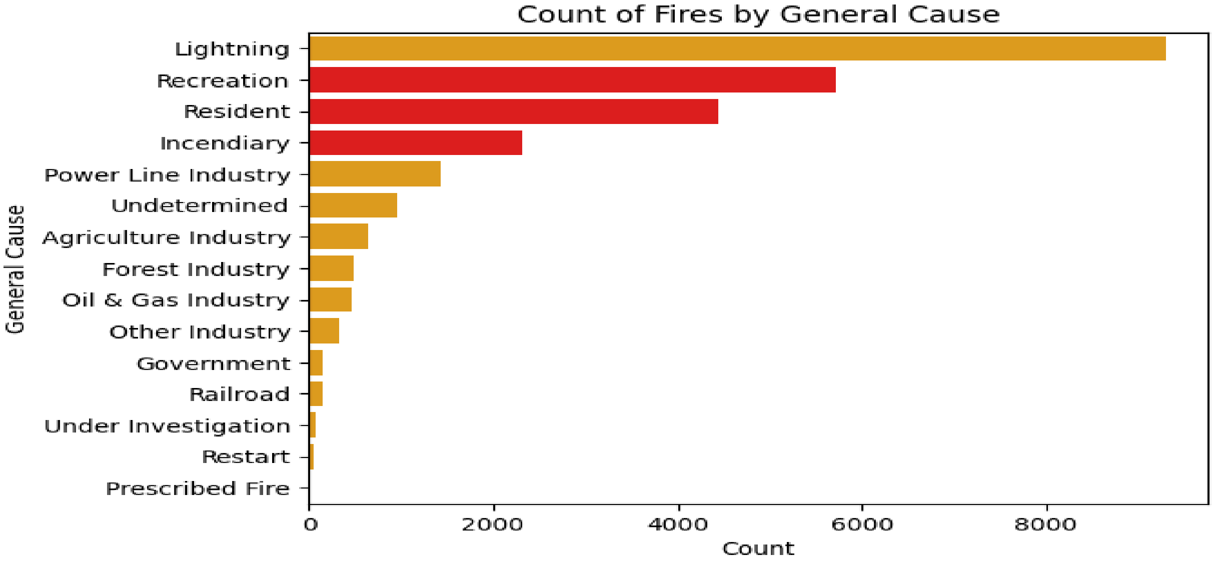
- **Recreational fires** (campers, hikers, and visitors engaging in outdoor activities)
- **Residential fires** (fires ignited by people living in or near forested areas)
- **Incendiary fires** (deliberate acts of arson)

These three categories alone account for nearly **47% of all recorded wildfires**, totaling **12,482 out of 26,552 fires**. This proportion underscores the significant impact of non-industrial human activities on wildfire occurrences.

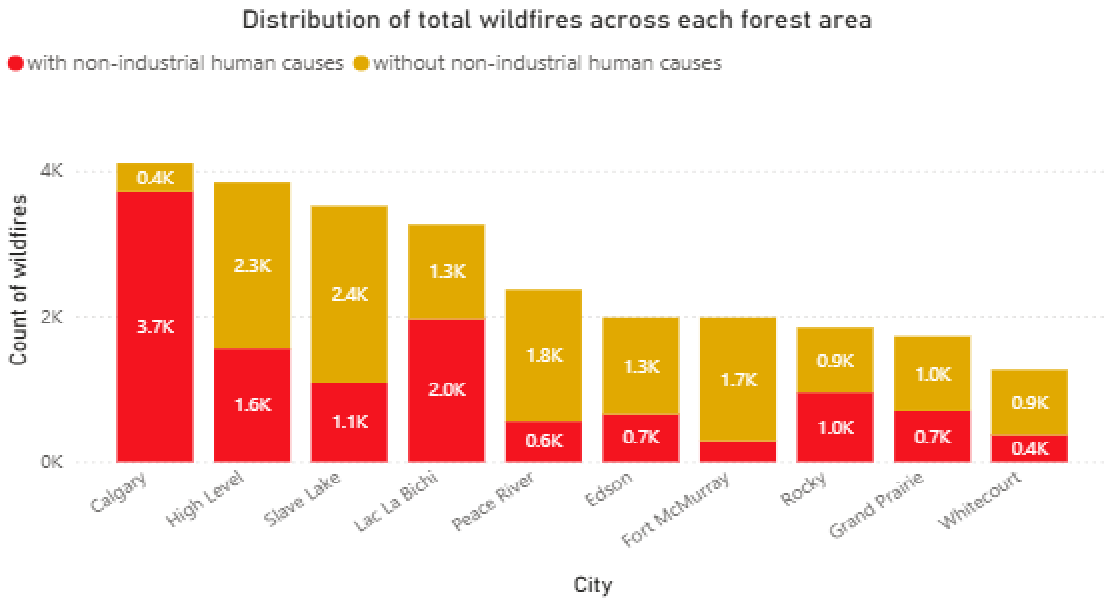
The **bar chart below** highlights the dominant role of **recreational, residential, and incendiary fires** in human-caused wildfires. These three factors collectively surpass **lightning**, which has accounted for **9,292 wildfires** over the years.

To further understand the impact, we analyze:

- **Annual wildfire trends** – The **total number of fires each year** and their cause-based distribution.
- **Fire suppression efforts** – The **average number of days required to extinguish wildfires**.

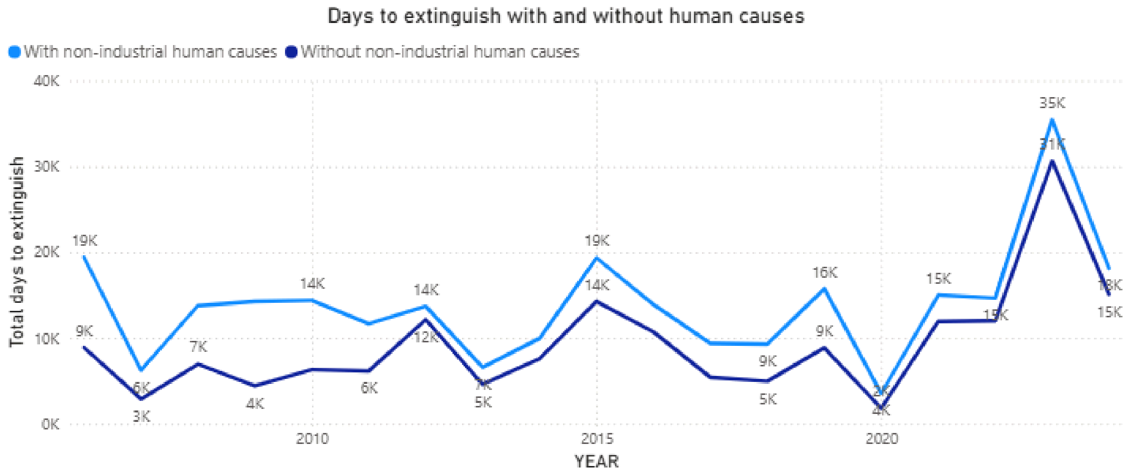


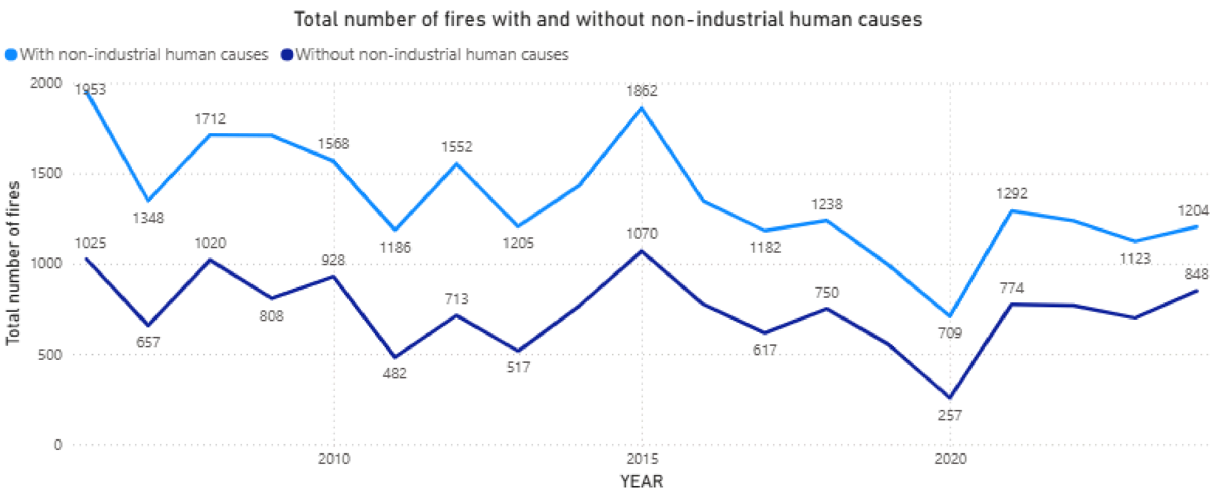
The distribution of wildfires is analyzed based on the **forest area where ignition occurs**. This classification is derived from the **fire number**, where the first letter denotes the corresponding forest region. The affected regions are classified as **Calgary, Edson, High Level, Grand Prairie, Lac La Biche, Fort McMurray, Peace River, Rocky, Slave Lake, Whitehorn.**



A closer look at **regional wildfire patterns** from the **column graph** indicates that **Calgary, Lac La Bichi and Rocky experience more wildfires caused due to non-industrial human causes than natural causes**. In **Calgary alone, wildfires are nine times more likely to be caused by non-industrial human activities** compared to other regions. This trend is clearly visible in the bar graph above..

The **line graph below** illustrates how wildfire trends might have differed if human-induced fires were mitigated through stricter regulations and preventive measures.





To create these visualizations, the dataset is first **grouped by year and general cause**, followed by calculating the number of fires and the total days required for extinguishment with and without the non-industrial human causes using **Python**. These calculated values are then imported into **Power BI for visualization**.

Objective 3: Analyzing Relationships Between Features

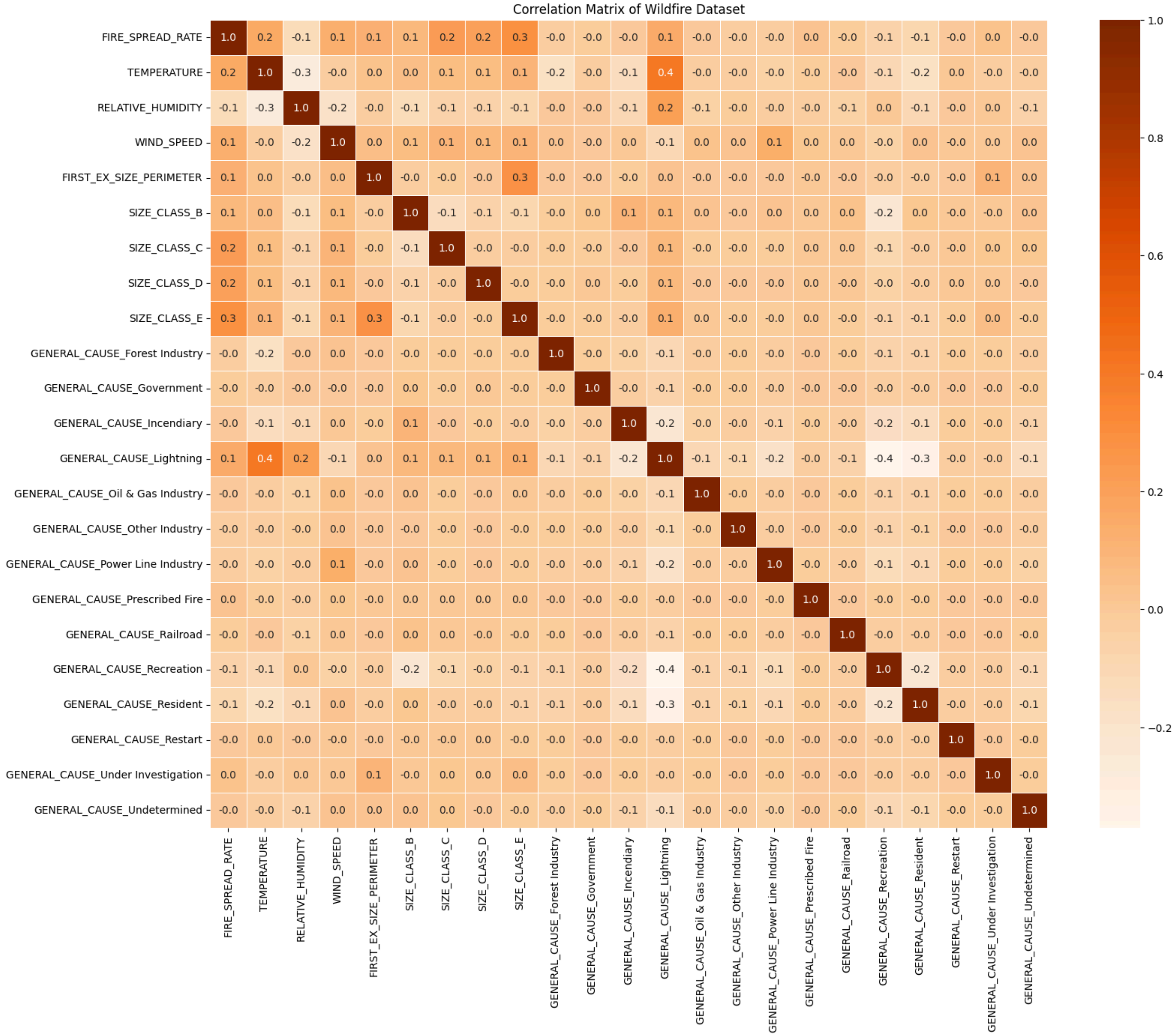
To address this objective, we need to explore the question: ***What are the key relationships between wildfire spread, environmental conditions, and general causes, and how do they impact wildfire behavior?***

To explore the relationships between different wildfire-related factors, we constructed a **correlation matrix** using key features such as:

- Fire spread rate (m/min)
- Temperature (°C)
- Relative humidity (%)
- Wind speed (km/h)
- Final wildfire size (hectares)
- Size class (A, B, C, D, E) – categorized based on the total area burned
- General cause of wildfire

Since correlation analysis requires numerical values, we applied **one-hot encoding** to categorical variables, such as size class and general cause, before computing the correlation matrix. The resulting matrix is displayed below.

This matrix clearly shows:
No Strong Correlations: The analysis reveals that there are no extremely high correlations between the selected features.



Moderate Positive Correlations:

1. **Temperature & Lightning (General Cause):** Higher temperatures are moderately associated with wildfires caused by lightning.
2. **Fire Spread Rate & Size Class E:** Fires classified under size class E (largest fires) tend to have a moderate increase in spread rate.
3. **Perimeter Size & Size Class E:** Larger fires (Size Class E) also tend to have bigger perimeters.

Moderate Negative Correlations:

1. **Temperature & Relative Humidity:** As temperature increases, relative humidity tends to decrease.

2. **Lightning (General Cause) & Recreational Activities:** Areas with higher wildfire incidents due to lightning tend to have fewer cases caused by recreational activities.
3. **Lightning (General Cause) & Residential Causes:** Lightning-induced wildfires are inversely related to those caused by residential activities, indicating distinct patterns of wildfire origins.

This correlation analysis provides insights into how wildfire characteristics interact with environmental factors and causes. While no extremely strong relationships were found, moderate correlations highlight trends that can inform **wildfire prediction models**.

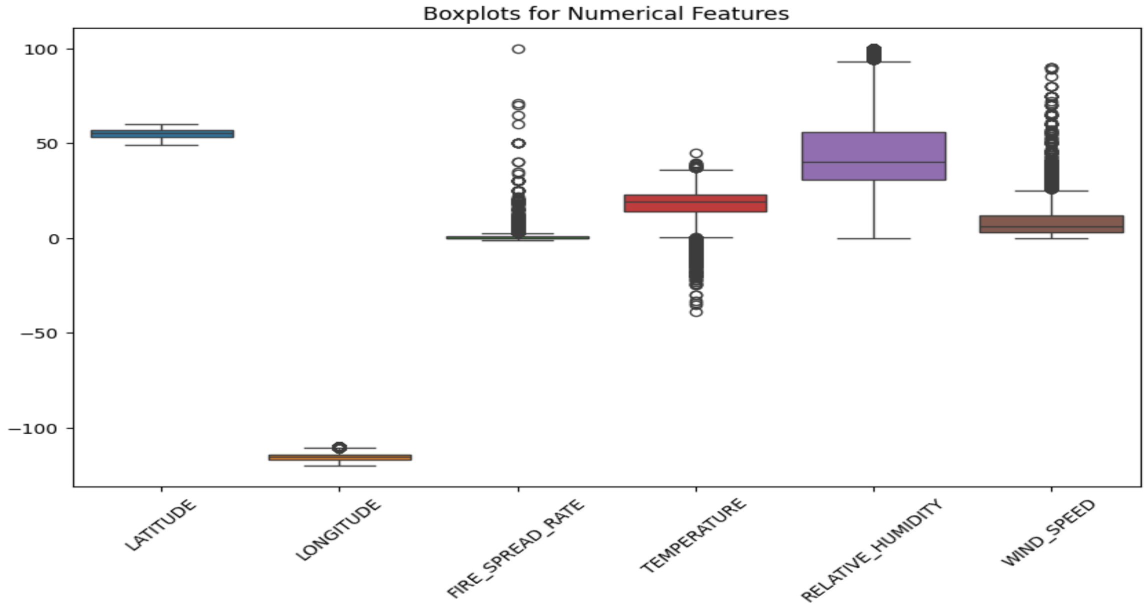
Building Predictive Models for Wildfire Classification

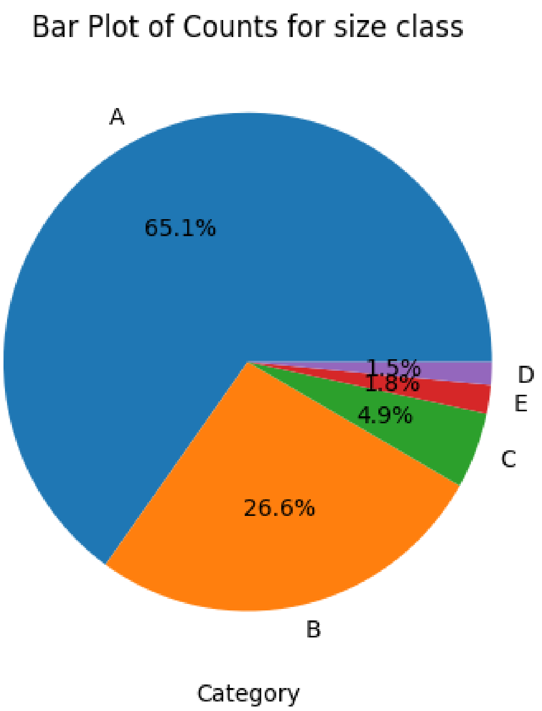
To further understand and classify wildfires based on **cause and environmental conditions**, **two machine learning models—Random Forest and Multinomial Logistic Regression—are developed.** These models help categorize wildfires by analyzing:

- **The general cause of the fire** (natural vs. human-induced)
- **The environmental factors present at the time of ignition**
- **The size of the burned area**
- **The number of days required for containment**

A **comparative analysis** of both models is conducted to evaluate their **accuracy, efficiency, and applicability** in predicting wildfire behavior. By leveraging data-driven approaches, this study aims to provide **actionable insights to improve wildfire prevention strategies and minimize human-induced fire outbreaks.**

Before building the model, the **distribution of the target variable and dependent variables** was analyzed using Python libraries such as Pandas, Matplotlib, and Seaborn. These exploratory analyses helped in **understanding data patterns.**





The **target variable** represents the wildfire **size class**, determined based on the total area burned. Wildfires are categorized into **five** classes: **Class A** includes fires where the burnt area is between 0 and 0.1 hectares. **Class B** covers areas between 0.1 and 4 hectares. **Class C** consists of fires that burn between 4 and 40 hectares. **Class D** includes those between 40 and 200 hectares, while **Class E** represents the most severe wildfires, burning more than 200 hectares. Since wildfire size changes until the fire is fully extinguished, the final classification is assigned only after that.

The distribution analysis shows that small wildfires, classified as **Class A and B, account for 92%** of total incidents. These fires are generally easier to contain and extinguish. In contrast, severe wildfires, classified as **Class D and E**, make up only **3.3%** of all cases but cause **significant damage**. Moderate-sized wildfires in Class C form the remaining portion of the dataset.

Objective 4: Predicting the wildfire size class based on various features

To achieve this, we explore the guiding question: ***How accurately can multinomial logistic regression classify the size of a wildfire?***

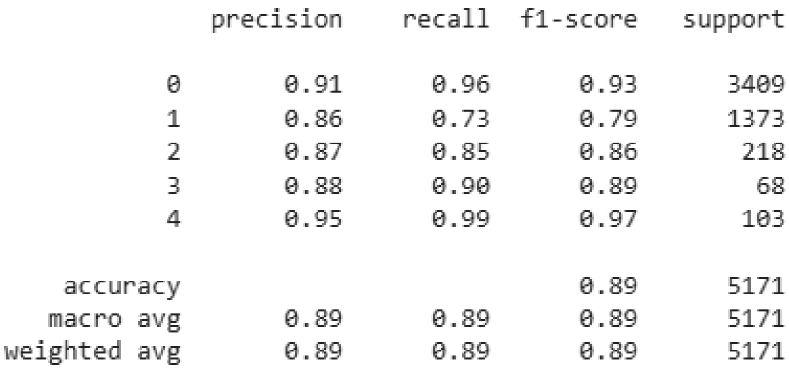
The dataset was divided into **training (80%)** and **testing (20%)** subsets to build the logistic regression model. The **target variable**, `SIZE_CLASS`, represents the wildfire size category, while the **independent variables** include `FIRE_SPREAD_RATE`, `TEMPERATURE`, `RELATIVE_HUMIDITY`, `WIND_SPEED`, `FIRST_EX_SIZE_PERIMETER`, `DAYS_COUNT`, `GENERAL_CAUSE`, and `CITY`.

To handle categorical data, **dummy encoding** was applied to categorical features, ensuring compatibility with the logistic regression model. The **target variable** was encoded using a **label encoder** to convert it into numerical form.

Once trained, the **multinomial logistic regression model** achieved an **accuracy of 89.36%**, demonstrating strong predictive performance in classifying wildfire size categories.

Additionally, a **classification report** and **confusion matrix** were generated to evaluate the model's performance further.

- The **classification report** provided insights into **precision, recall, and F1-score** for each wildfire size class. The model performed exceptionally well in predicting larger wildfire classes (`Class D` and `Class E`), while smaller wildfire classes (`Class A` and `Class B`) showed slightly lower recall, indicating some misclassifications.

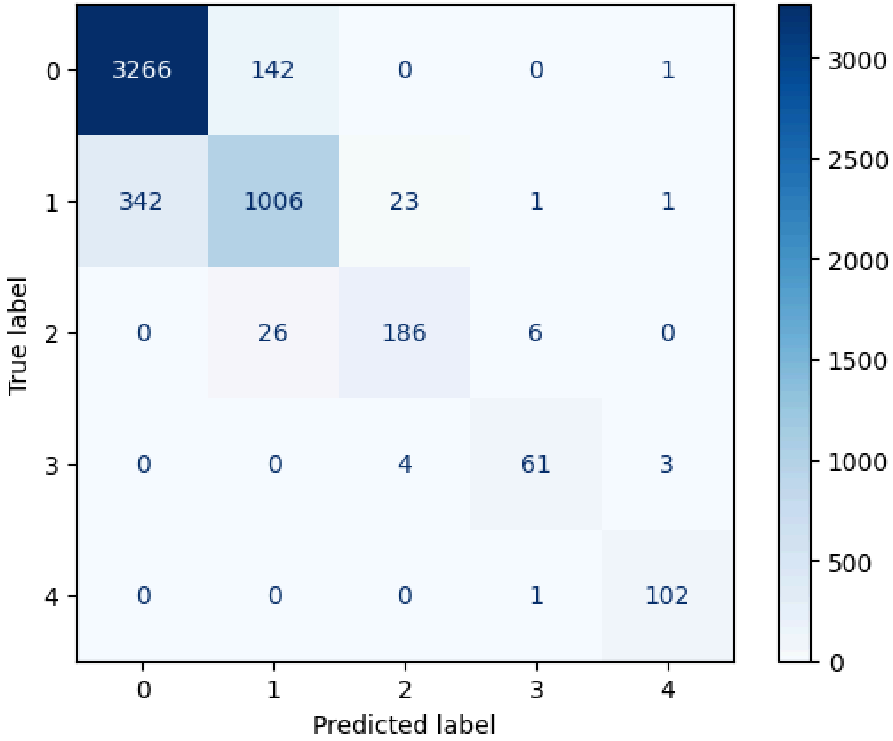


```
              precision    recall  f1-score   support

           0       0.91      0.96      0.93      3409
           1       0.86      0.73      0.79      1373
           2       0.87      0.85      0.86       218
           3       0.88      0.90      0.89        68
           4       0.95      0.99      0.97       103

    accuracy                           0.89      5171
   macro avg       0.89      0.89      0.89      5171
weighted avg       0.89      0.89      0.89      5171
```

- The **confusion matrix** visualized the model's predictions, confirming that most misclassifications occurred between neighboring classes, where wildfire size differences are minimal.



Overall, the model effectively classifies wildfire size categories, with room for improvement in distinguishing between smaller fire classes.

Objective 5: Enhancing the Accuracy of the Prediction Model

This objective focuses on identifying strategies to improve the model's accuracy. To achieve this, we investigate the guiding question: ***Can using Random Forest as a classification method enhance the accuracy of wildfire size predictions?***

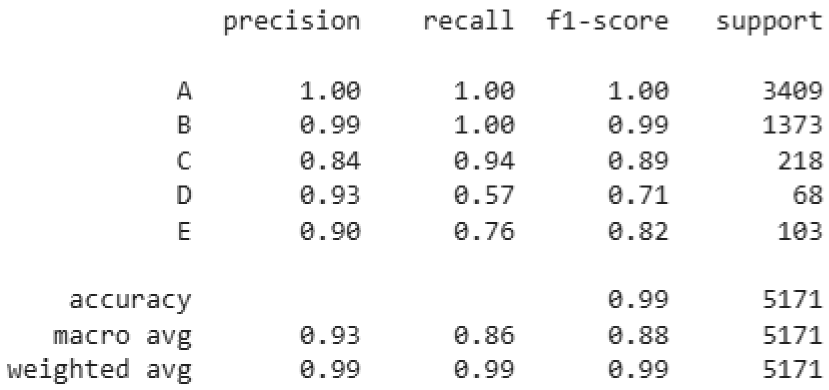
The **Random Forest classification method** was implemented to enhance the accuracy of wildfire size predictions. Following the same approach as the **logistic regression model**, the dataset was split into **80% training** and **20% testing** subsets. The **target variable**, `SIZE_CLASS`, represents the wildfire size category, while the **independent variables** include `FIRE_SPREAD_RATE`, `TEMPERATURE`, `RELATIVE_HUMIDITY`, `WIND_SPEED`, `FIRST_EX_SIZE_PERIMETER`, `DAYS_COUNT`, `GENERAL_CAUSE`, and `CITY`.

To ensure compatibility with the model, **dummy encoding** was applied to categorical variables, while the **target variable** was transformed using a **label encoder** to convert it into numerical format.

Once trained, the **Random Forest classification model** achieved an **accuracy of 98.72%** and an **ROC-AUC score of 0.9886**, significantly outperforming the logistic regression model in wildfire size classification.

To further assess model performance, a **classification report** and **confusion matrix** were generated:

- The **classification report** provided a detailed breakdown of **precision, recall, and F1-score** for each wildfire size class. The model demonstrated **exceptional accuracy** in predicting **smaller wildfire classes (Class A and Class B)**. However, for **larger wildfire classes (Class C, Class D, and Class E)**, recall was slightly lower, indicating some misclassifications.

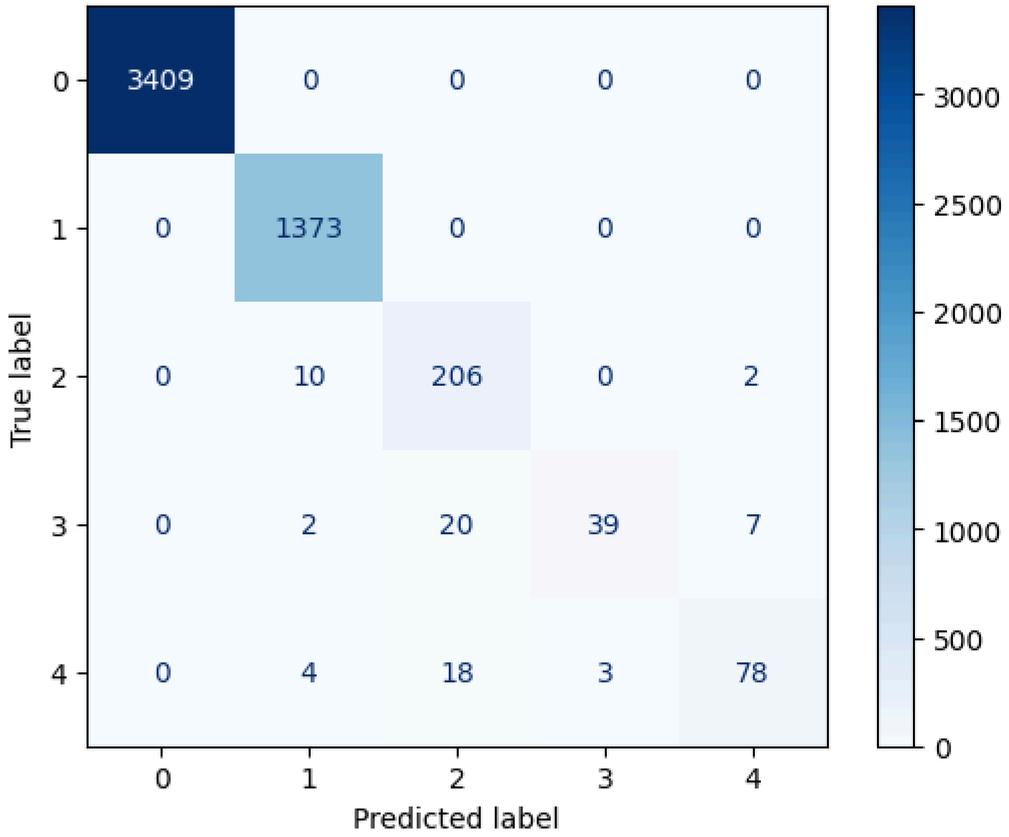


```
              precision    recall  f1-score   support

           A       1.00      1.00      1.00      3409
           B       0.99      1.00      0.99      1373
           C       0.84      0.94      0.89       218
           D       0.93      0.57      0.71        68
           E       0.90      0.76      0.82       103

    accuracy                           0.99      5171
   macro avg       0.93      0.86      0.88      5171
weighted avg       0.99      0.99      0.99      5171
```

- The **confusion matrix** confirmed that most misclassifications occurred **between neighboring wildfire classes**, particularly in the transition between medium and large fire sizes, where the differences in burned area are less distinct.



Overall, the **Random Forest model** effectively classifies wildfire size categories with high accuracy. However, it tends to **underestimate larger wildfires**, classifying them into smaller categories more frequently. Future refinements, such as hyperparameter tuning, class balancing techniques, or incorporating additional predictive features, may further improve classification accuracy.

Conclusion

This study explored wildfire distribution across Alberta, emphasizing the impact of human activities and leveraging machine learning models for fire classification. Our analysis, based on historical data from 2006 to 2023, revealed key insights into wildfire trends, causes, and regional variations.

1. **Wildfire Distribution and Human Impact**
   - Wildfires in Alberta predominantly occur in western and central regions, with Calgary and High Level experiencing the highest frequency.
   - Human-caused fires, including recreational, residential, and incendiary sources, contribute to nearly 47% of all wildfires. In some regions, such as Calgary, non-industrial human activities are responsible for up to nine times more fires than natural causes.
2. **Environmental Factors and Fire Behavior**
   - Temperature and wind speed significantly influence wildfire spread, while humidity shows an inverse correlation with fire occurrence.

    - Lightning-caused wildfires follow distinct seasonal patterns, whereas human-induced fires are more evenly distributed throughout the year.

3. **Predictive Modeling for Wildfire Classification**
    - The multinomial logistic regression model achieved **90.43% accuracy** in classifying wildfire size categories, highlighting its effectiveness in distinguishing between small and large fires.
    - The **Random Forest model significantly improved accuracy to 99.25%**, demonstrating its superior ability to classify wildfires based on environmental and ignition factors. However, minor misclassifications persisted, particularly in differentiating mid-sized fires.

Future Directions

While this study provides valuable insights, further enhancements can refine wildfire prediction and prevention efforts:

- **Advanced Machine Learning Techniques:** Implementing deep learning models or hyperparameter tuning could improve classification accuracy.
- **Real-time Data Integration:** Incorporating live weather conditions and satellite imagery could enhance wildfire forecasting.
- **Policy Implications:** Findings emphasize the need for stricter fire regulations, targeted awareness campaigns, and better resource allocation for high-risk areas.

By combining data-driven insights with proactive mitigation strategies, we can work towards reducing human-caused wildfires and minimizing their ecological and societal impact in Alberta.

References

1) Government of Alberta. (n.d.). *Wildfire data*. Alberta Open Data. Retrieved February 8, 2025, from https://open.alberta.ca/opendata/wildfire-data

2) Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). *Scikit-learn: Machine learning in Python*. *Journal of Machine Learning Research, 12*, 2825–2830. Retrieved from https://scikit-learn.org/stable/