Forest Fires Prediction and Prevention

Ashritha Kumari B, Dhrumil Shah, Madhulika Datta,
Prudhvi Chowdary Chirumamilla, Shashank Reddy

Agenda

01 Industry Case Study

02 Significant Paper

03 Project Presentation

Industry Case Study

Introduction and Importance of Forest Fire Prevention

- Forest fires have devastating impacts. They endanger lives, destroy ecosystems, and result in substantial economic losses. In the United States alone, forest fires have caused approximately 16 billion dollars of ecological damages. Our presentation focuses on the critical area of forest fire prediction and prevention."
- Preventing forest fires is vital for several reasons. It helps protect natural habitats, biodiversity, and human safety. Additionally, it saves billions of dollars in firefighting efforts and property damage.
- Our research draws information from multiple sources, including medium blogs, company white papers (Descartes Lab, Orbital Insights, International Association of Wildland Fire), TedTalks, and interviews.

Industry Case Study

Data Collection and Preprocessing

- Geospatial data from satellite sources like MODIS, Landsat, and Sentinel 1, 2, and 3 is collected for tracking forest fire locations and behavior.
- Meteorological factors such as temperature, humidity, wind speed, and precipitation are gathered from satellites and organizations like Accuweather to understand their influence on fire behavior.
- Geographic data, including topography, soil types, land cover, and vegetation density, is integrated to assess fire risk, along with historical data for identifying trends and patterns in fire locations and causes.
- Because of the vast volume of data collected, it must be digested and processed in near real-time, which necessitates the deployment of a cloud-based supercomputer. Companies use complex tools like Python, Cloud Technologies, Panoply, Climate Data Operators (CDO) and GDAL (Geospatial Data Abstraction Library) for data format conversion,caliberation, correction, subsetting, and transformation.

Industry Case Study

Feature Engineering

- Multispectral satellite data is utilized to extract information about Earth's surface, including vegetation, water content, and mineral composition.
- Methodologies like the Normalized Burn Ratio (NBR) are employed to develop burn area signatures and measure the extent and severity of burns.
- Geographic Information System (GIS) tools analyze data related to topography, land cover, soil types, and terrain characteristics, which is then combined with other variables to create features that aid in assessing fire risk. Principal Component Analysis (PCA) is used for dimensionality reduction when dealing with high-dimensional data.

Industry Case Study

Model Training and Optimization

- Machine learning methods, including K-Nearest Neighbors, Decision Trees, Gradient Boosting, and Random Forest, are used for wildfire detection and monitoring.
- These algorithms classify locations based on historical fire-prone areas, identify key factors affecting forest fires, and predict fire spread while prioritizing management areas.
- In addition, they leverage Wireless Sensor Networks and various machine learning techniques like Backpropagation Network, Radial Base Function Network, Dynamic Learning Vector Quantization, and Multilayer Perceptron to enhance forest fire prediction and prevention.
- In forest fire prediction, optimizing machine learning algorithm hyperparameters is crucial. Ensemble models like Random Forest benefit from parameter tuning, which can significantly impact model convergence speed and prediction accuracy. One optimization technique, Particle Swarm Optimization (PSO), is employed to fine-tune key parameters within Random Forest models, resulting in enhanced accuracy and generalizability for assessing forest fire risk.

Industry Case Study

Evaluation

- They use several evaluation metric techniques such as Accuracy, Precision, Recall, F1 Score, AUC-ROC, Confusion Matrix and K-fold Cross Validation Techniques.
- They work on automated wildfire monitoring and burn scar evaluation, to demonstrate that their methods are effective in giving significant insights for post-fire assessments and recovery planning.

Industry Case Study

Deployment Strategies

In the deployment phase, several strategies are employed to enhance forest fire prevention and management:

- Alarm Systems and Public Awareness: Automated alarm systems are developed to issue warnings when the risk of fire is high, ensuring timely response.
- Collaboration and Coordination: The National Interagency Fire Center (NIFC) acts as a vital wildfire support center, fostering collaboration among various governmental and private entities.
- Monthly Seasonal Outlook Podcast: The NIFC produces a monthly seasonal outlook podcast, offering valuable insights into future wildfire risks.
- NIFC and IAWF advocate the use of controlled fires as a proactive method to prevent forest fires. They recognize that historical forest landscapes have changed significantly due to factors like fire suppression, logging, and tree removal.

Industry Case Study

Success in Prevention

- With the rise of several issues like Global Warming and climate changes the forest fire spread tends to increase year by year. However, on seeing the reports, we found these agencies were able to save 98% of the 66,255 forest fires spreading in the US which had the capacity to burn over 75,34,403 acres and also were able to successfully prevent 70 fires from burning further.

Significant Paper

Spatial Prediction of Wildfire Susceptibility Using Field Survey GPS Data and Machine Learning Approaches

Ghorbanzadeh, O.; Valizadeh Kamran, K.;Blaschke, T.; Aryal,J.; Naboureh, A.; Einali, J.; Bian, J. Spatial Prediction ofWildfire SusceptibilityUsing Field Survey GPS Dataand Machine Learning Approaches. Fire 2019, 2, 43. https://doi.org/10.3390/fire2030043

Significant Paper

Introduction

- The increasing availability of free remotely sensed data has enabled the precise locations of wildfires to be reliably monitored.
- Global warming and climate change causes dryness and long-term water scarcity which causes forest fires. Forest fires in return increase carbon dioxide in the atmosphere while they burn.
- Data used is for a place in Northern Iran named Amol County, the data spans from 2012 to 2017.
- Dataset has five ecological factors namely: topographic, meteorological, vegetation, anthropological and hydrological factors.

Significant Paper

Research Related to this Problem

- Piyush Jain and Mike D. Flannigan discusses the history and evolution of applying AI in wildfire science and management.
- Younes Oulad Sayada’s paper demonstrates a practical application of Artificial Neural Networks and SVM, in monitoring and predicting wildfires based on satellite data.
- Ethan Weber and Antonio Torralba focused on using social media posts and image data for disaster understanding and introduces the Incidents1M Dataset.
- M. B. Joseph and M. W. Rossi integrates a 30-year wildfire record with meteorological and housing data in spatiotemporal Bayesian statistical models to predict wildfire extremes.

Significant Paper

Proposed Methodology

- Preparing the conditioning factors based on five main factors, namely topographic,meteorological, anthropological, vegetation and hydrological.
- •Using a four-fold Cross validation and dividing the merged dataset into four different equal-sized folds.
- Classifying the wildfire susceptibility to different label, from very low to very high.
- Applying ANN, SVM and RF to predict the wildfire susceptibility.
- Validating performances of each ML approach using the ROC curve.

Significant Paper

Experimental Results

Accuracy was calculated for all three models and different CV.

ML	AUC-Fold1	AUC-Fold2	AUC-Fold3	AUC-Fold4	Cross-Validation (CV)
ANN	0.74	0.71	0.73	0.79	0.74
SVM	0.78	0.78	0.82	0.75	0.79
RF	0.89	0.85	0.94	0.85	0.88

Significant Paper

Conclusion and Future Scope

- This study trained on MODIS hotspots dataset and factors that contributed to wildfires from 2012 to 2017.
- The performance of ML approaches, ANN, SVM and RF, were evaluated by ROC curve.
- It was found that central, east, southern and northern regions of the study area were more susceptible to wildfires.
- The proposed workflow can easily be applied to other fire-prone regions like California, Australia and Spain.

Significant Paper

Relation to Project

- We have used the conditioning factors such as topographic, meteorological, anthropological, vegetation and hydrological.
- Similar to their ML approaches, we have used SVM and RF algorithms for prediction susceptibility.
- Validating the performance of each model using ROC curve.

Significant Paper

Key Learnings

- Understanding the various factors that contribute to the occurrence and spread of wildfires.
- Recognizing relationships between factors and wildfire risk that impact wildfire susceptibility aids model performance.
- mportance and use of cross-validation techniques in evaluating the performance and generalizability of machine learning approaches.
- Knowledge of the ROC curve as an evaluation metric for measuring the accuracy and performance of classification algorithms.
- Familiarity with different classification algorithms, such as SVM and RF

Significant Paper

Appendices

- Significance of the paper, Publisher : MDP, Journal : Fire
- Summarized key learnings are substantial.
- Topics and contents of the paper are relevant to the course work.
- The presentation was impactful and answered the questions confidently and correctly.
- The project is built upon the significant paper.
- Quality of the slides are concise, well-organized, follow clear format ensuring readability.

Wildfire Size Forecasting: A Machine Learning Approach

Ashritha Kumari B, Dhrumil Shah, Madhulika Datta, Prudhvi Chowdary Chirumamilla, Shashank Reddy

Abstract

- Developed a comprehensive forestfire prediction system using machine learning algorithms, including SVM, Naive Bayesian, DecisionTree, Random Forest, and Logistic Regression.
- Analyzed a dataset encompassing factors like wind, humidity, temperature, and vegetation to predict the severity of wildfires.
- Evaluated the performance of different classifiers and found Random Forest to be the best- performing model with an accuracy of 75.44%.
- Demonstrated the importance of machine learning in forest fire prediction and contributed to Sustainable Development Goals by developing a model capable of predicting forest fire size based on various environmental factors.

Introduction

- Forest fires pose a significant threat to ecosystems, property, and human life.
  - Devastating ecological impacts, including habitat destruction and biodiversity loss
  - Substantial financial costs due to property damage and firefighting expenses
- Preventing forest fires is essential for safeguarding ecosystems, property, and human life.
  - Implementing proactive measures to reducefirerisks
  - Educating the public on fire prevention practices
  - Enforcing strict regulations on activities that could ignite fires
- Advanced technologies and machine learning algorithms can enhance forest fire prediction.
  - Utilizing data on weather conditions, vegetation, and other factors
  - Employing algorithms like SVM, Naive Bayesian, Decision Tree, Random Forest, and Logistic Regression
- Accurate prediction models can aid in forest fire prevention and containment.
  - Classifying fire severity at an early stage
  - Developing effective mitigation strategies
  - Adapting to climate change-induced fire risks

Objective

This project aims at developing a classification model that is capable of predicting the size of forest fires based on vegetation factors as well as meteorological factors, which is an important contribution to achieving the Sustainable Development Goals.

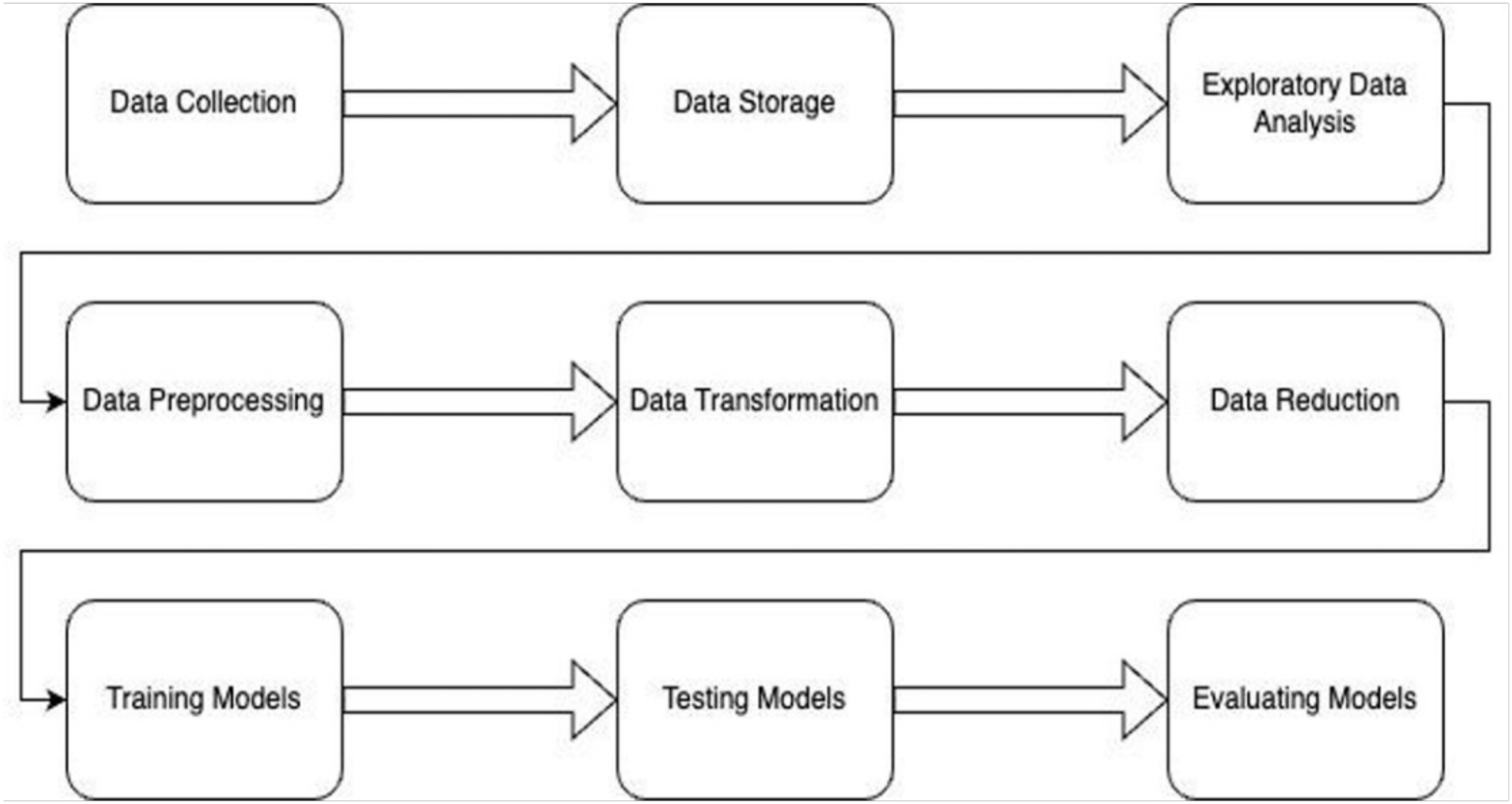
- The target class was initially imbalanced when the fire size class consisted of A-G classes, combined A-B into one called “0” and C-G into one called “1”.
- Therefore, 0 idicates small fire <25 Acres and 1 represents a widespread fire >25Acres.

Fire Size Class	**Acres**
A	<= 0.25 Acres
B	0.26 - 9.9 Acres
C	10-99.99 Acres
D	100-299 Acres
E	300-999 Acres
F	1000-4999 Acres
G	5000+ Acres

Sustainability

- Wildfires can damage agricultural land and forest resources, leading to food insecurity and loss of livelihoods for rural communities, and thereby impacting **SDG2: Zero Hunger**.
- Wildfires can release harmful pollutants into the air, posing health risks to people in the surrounding area, and thereby impacting **SDG3: Good Health and Well-being.**
- Wildfires can damage infrastructure such as power lines and homes, requiring costly repairs, and thereby impacting **SDG9: Industry, Innovation and Infrastructure.**
- Wildfires can destroy habitats and biodiversity, affecting ecosystems and the livelihoods of communities that depend on them, and thereby impacting **SDG15: Life on Land.**
- Wildfires release greenhouse gasses and contribute to global heating, accelerating climate change, and thereby impacting **SDG13: Climate Action.**

Methodology



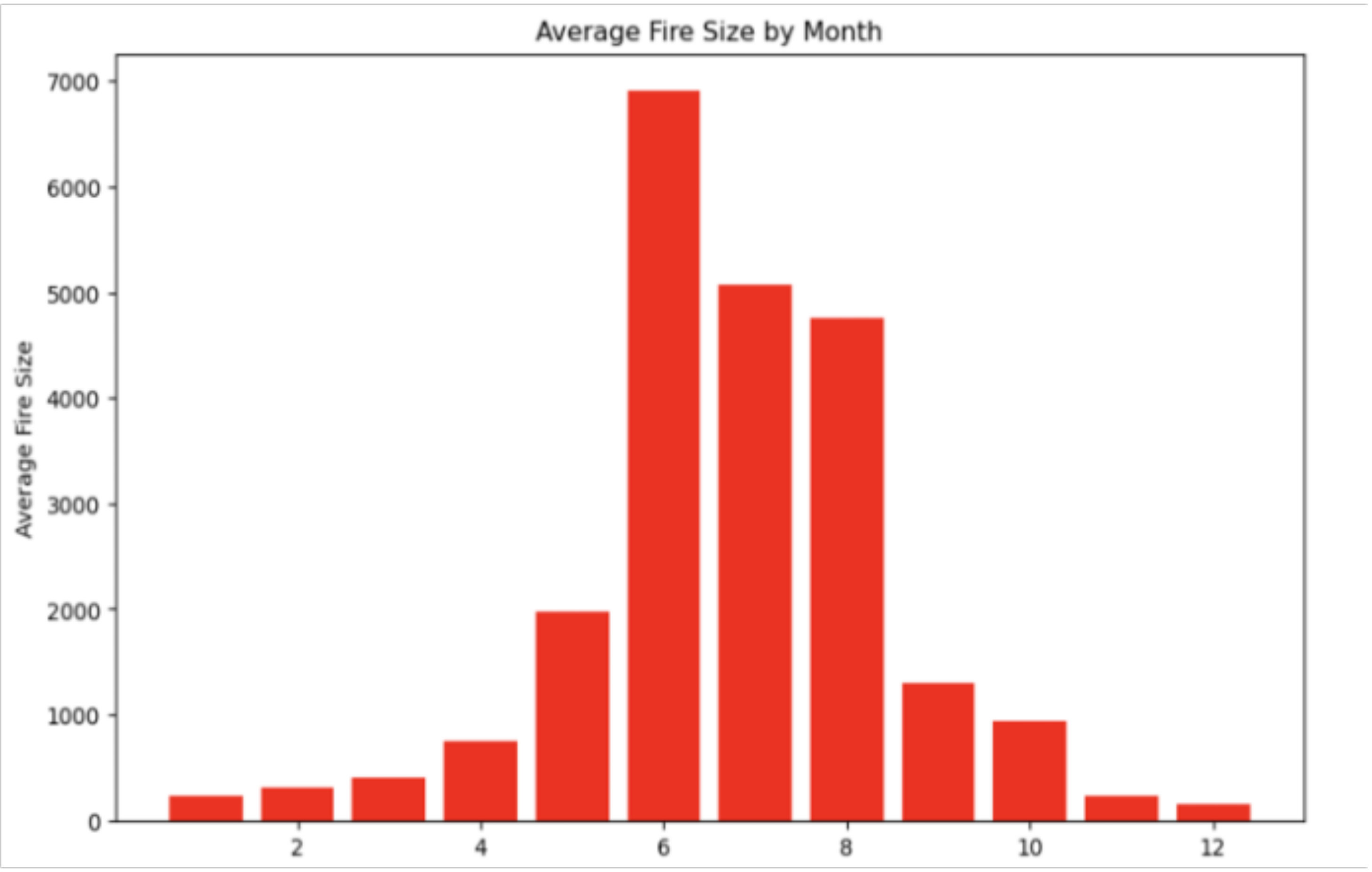
Data Collection and Storing

The dataset utilized in this project is a subset of a larger collection of data encompassing 1.88 million fires in theUnited States and ranges from 1991 - 2015. This subset was created by combining historical weather and vegetation data sharing the same latitude and longitude as 50,000 randomly selected fire samples.

- Fire_size_class - Class of Fire Size (A-G)
- Stat_cause_descr - Cause of Fire
- Latitude - Latitude of Fire
- Longitude - Longitude of Fire
- Discovery_month - Month in which Fire was discovered
- Vegetation - Dominant vegetation in the areas (can save some factors of vegetation)
- Temp_pre - temperature in deg C at the location of fire up to 30, 15 and 7 days prior
- Temp_cont - temperature in deg C at the location of fire up to day the fire was
- Wind_pre - wind in deg C at the location of fire up to 30, 15 and 7 days prior
- Wind_cont - wind in deg C at the location of fire up to day the fire was
- Prec_pre - Precipitation in deg C at the location of fire up to 30, 15 and 7 days prior
- Prec_cont - Precipitation in deg C at the location of fire up to day the fire was
- Hum_pre - Humidity in deg C at the location of fire up to 30, 15 and 7 days prior
- Hum_cont - Humidity in deg C at the location of fire up to day the fire was
- Remoteness - non-dimensional distance to closest city

Exploratory Data Analysis (EDA)

- Exploratory data analysis (EDA) is an iterative process that involves summarizing and analyzing datasets in order to uncover potential problems, better understand the underlying structure and patterns of the data, and produce hypotheses for additional research.
- Examples: Barchart of month-wise avg fire_size and choropleth map of distribution of fire_size class.



Data Preprocessing I

Data Cleaning:

- Dropped columns with null values and irrelevant information: Three columns were dropped because they had null values and were not important for the analysis.
- Removed irrelevant columns: Columns such as state, putout_time, and disc_pre_year were removed because they were not relevant to the analysis.
- Handled missing values: Missing values were handled by dropping rows with missing values in the wind, temperature, humidity, and precipitation columns.

Data Preprocessing II

Data Transformation:

- Scaled data using MinmaxScaler: The MinmaxScaler was used to scale the range of wind, temperature, humidity, and precipitation columns to be between 0 and 1.
- Clubbed classes for target variable: The target variable, which was multiclass, was imbalanced, so we clubbed the classes into two categories: fire size less than 9.9 acres and fire size greater than 9.9 acres.
- 3.Used target encoding for categorical variables: Target encoding was used to convert categorical variables into numerical values based on the mean of the target variable corresponding to each category. This technique helps capture correlations between categorical features and the target variable in a supervised learning environment.

Model Selection

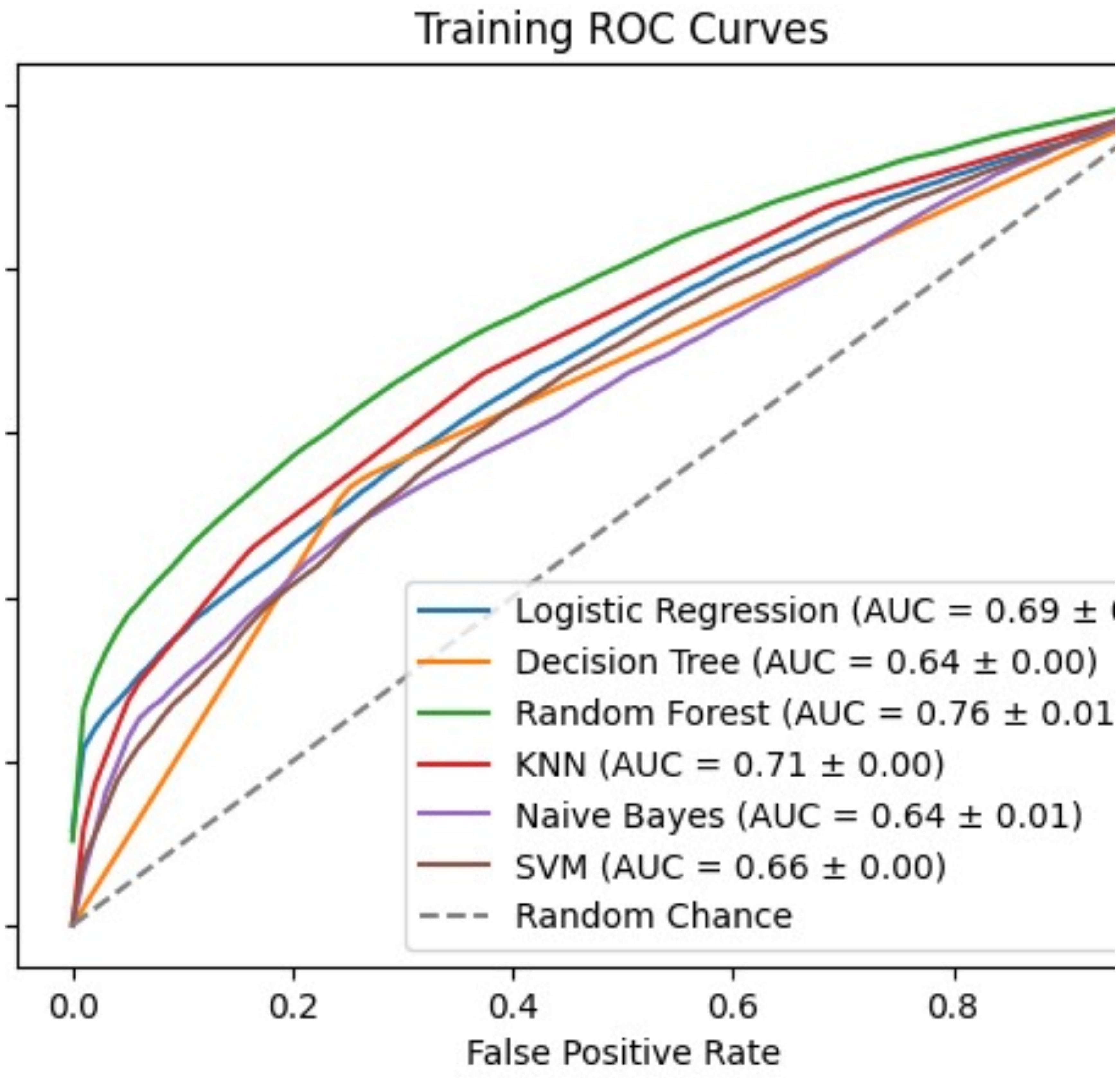
Accurately predicting forest fires is crucial for preventing environmental damage and safeguarding human lives. Machine learning algorithms have emerged as powerful tools for forest fire prediction, offering promising results in various studies. Among the numerous models available, we chose to go forward with:

- SVM
- Decision Tree
- Random Forest
- KNN
- Logistic Regression
- Naive Bayes

Model Validation

Model validation is crucial to ensure the reliability and generalizability of machine learning models for forest fire prediction.

- Techniques such as StratifiedKFold can be used to ensure a balanced representation of target classes in both the training and testing sets, preventing biases and ensuring fairness in the model's performance.
- StratifiedKFold repeatedly divides the data into k folds while maintaining the same proportion of target classes in each fold, providing a more comprehensive assessment of the model's generalizability.
- The model is trained on each fold except one, and its performance is evaluated on the excluded fold, allowing for a more accurate assessment of its predictive power.
- In this study, we employed 5-fold cross validation to ensure a robust evaluation of our machine learning models' performance.



Model Evaluation

- Evaluating the performance of machine learning models is crucial for ensuring their effectiveness in real-world applications.
- Used several evaluation methods namely, Accuracy, Precision, Recall, F1 score, AUC/ROC curve.
- Upon comparing the models, Random Forest turned out to be the best Classifier with Accuracy of 75.44%, recall of 75.18% and precision of 74.64%.

Model Comparision

Model	Accuracy	Precision	Recall	F1 Score
LR	72.99	71.75	72.25	68.81
DT	68.35	68.11	67.95	68.02
RF	75.44	74.64	75.18	73.34
KNN	71.18	69.73	70.95	69.88
NB	67.70	66.15	67.81	66.43
SVM	67.50	74.79	67.48	56.07

Conclusion

- Random Forest was the best performing classifier for predicting forest fire size, with an accuracy of 75.44%, a recall of 75.18%, and a precision of 74.64%.
- KNN and Logistic Regression were the second best performing models, also capable of efficiently classifying the fire_size class.
- The output from these models can be used to aid in research and help authorities concerned minimize fire spread by creating public awareness, coordinating, and collaborating with several organizations.

Questions?

Thank You