PREDICTING BURNED AREA OF FOREST FIRES USING METEOROLOGICAL DATA

A Study by Taylor Callahan

TABLE OF CONTENTS

01 INTRODUCTION

02 BACKGROUND

03 DATA

04 METHODS

05 LIMITATIONS

06 CONCLUSION

01

INTRODUCTION

PROBLEM AND OBJECTIVE

BUSINESS PROBLEM

The increasing occurrence, spread, and impact of forest fires presents a pressing problem in the world today. In Portugal, a country heavily affected by forest fires, over 2.7 million hectares of forest area were destroyed from 1980 to 2005. This analysis aims to examine meteorological factors in conjunction with burned area data from forest fires in Portugal.

OBJECTIVE

The objective is to create predictions for the impact of forest fires and identify high-risk areas. This analysis holds the potential to enhance early detection, support fire management decisions, and effectively warn the public of fire risk.

02

BACKGROUND

HISTORY AND RESEARCH SURROUNDING FOREST FIRES

CURRENT TRENDS

There has been a notable increase in their occurrence, spread, and impact, affecting both human and natural ecosystems. Accurately predicting the extent of forest fire damage remains a challenge

PRIOR RESEARCH

Large and severe forest fires are associated with warm and dry conditions, which occur with a warming climate. The role of meteorological factors has often been overlooked.

FUTURE PROJECTIONS

Based on projections, current trends, and simulation modeling, warmer and drier conditions will lead to longer fire seasons, increasing the frequency and extent of fires

03

DATA

DATA SOURCE

- The dataset used in this analysis is derived from "A Data Mining Approach to Predict Forest Fires using Meteorological Data" by P. Cortez and A. Morais
- The dataset captures information from the Montesinho natural park in Portugal, spanning from January 2000 to December 2003
- The dataset comprises 517 observations.

DATA DICTIONARY

Variable	Description
X	X-axis spatial coordinate within the Montesinho park map
Y	Y-axis spatial coordinate within the Montesinho park map
Month	Month of the year (jan to dec)
Day	Day of the week (mon to sun)
FFMC	Fine Fuel Moisture Code from the FWI system
DMC	Duff Moisture Code from the FWI system
DC	Drought Code from the FWI system
ISI	Initial Spread Index from the FWI system
Temp	Temperature in Celsius degrees
RH	Relative humidity in percentage
Wind	Wind speed in km/h
Rain	Rainfall in mm/m^2
Area	Burned area of the frest (in hectares)

DATA PREPARATION

CONVERT TO NUMERIC

The variables X and Y were converted to numeric format to represent the spatial coordinates accurately

CONVERT TO FACTOR

The variables month and day were converted to factors to categorize them appropriately for analysis

OVERSAMPLING

Non-zero instances were duplicated, effectively increasing the representation of non-zero area cases in the data. This addressed the issue of imbalanced data.

TRANSFORMATION

To address the challenge of missing values during resampling, a constant value of 1 was added to all numeric variables.

04

METHODS

EXPLORATORY DATA ANALYSIS

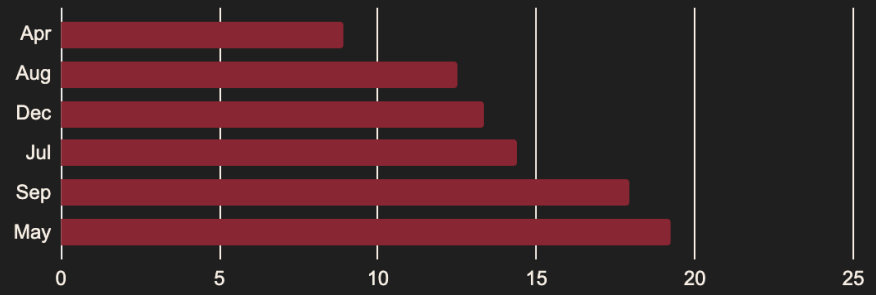
BURNED AREA IS RIGHT SKEWED

LOW CORRELATION WITH AREA

NO CLEAR MONTHLY TREND

MONTHS WITH LARGEST BURNED AREA

April	**8.89**	July	**14.37**
August	**12.49**	September	**17.94**
December	**13.33**	May	**19.24**



MODELING

PREPARATION

The dataset was split into training and testing sets. The hyperparameters were tuned to optimize performance.

EVALUATION

RMSE and R-Squared were chosen as evaluation metrics.

RANDOM FOREST

Known for its robustness against overfitting and ability to handle interactions between variables.

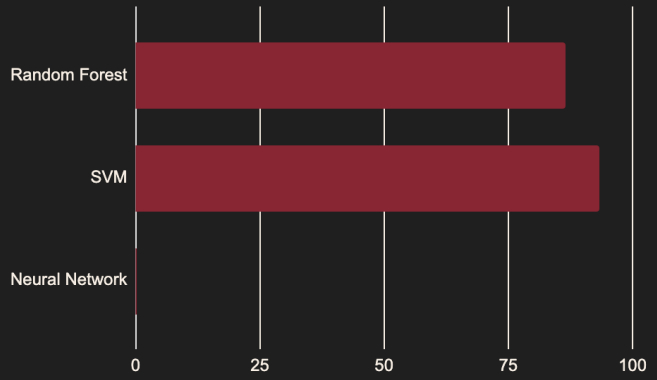
SVM

Ability to find optimal hyperplanes in high-dimensional spaces

NEURAL NETWORK

Excels in learning intricate patterns and relationships

RESULTS



86.6% **93.4%** **0.30%**

of the variance in burned area can be explained by the model.

05

LIMITATIONS

LIMITATIONS AND CHALLENGES

SCOPE

The dataset is limited to a specific time period and geographical location. This narrow scope may limit the generalizability of the findings to other regions or time periods.

AVAILABLE VARIABLES

The dataset does not include information on potential factors that may influence forest fires, such as human activities (e.g., arson) or topographical features (e.g., slope, elevation).

IMBALANCED DATA

A majority of the instances have no or minimal burned area. This can pose challenges in modeling and accurately predicting instances with significant burned areas.

06

CONCLUSION

MODEL SELECTION AND PRACTICAL USE

MODEL SELECTION

SVM emerged as the best-performing model, exhibiting the highest R-Squared value and the lowest RMSE. Therefore, SVM was identified as the most suitable model for predicting the extent of forest fire damage based on the considered evaluation metrics.

PRACTICAL USE

There is a potential to enhance early detection systems, support fire management decision-making processes, and effectively communicate fire risk to the public. Stakeholders can take proactive measures to mitigate damage caused by fires and protect human and natural ecosystems.

THANK YOU

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**