Project Report
Forest Fire Prediction

Project Title: Forest Fires Prediction
Name: Mihir Panchal
Date: October 2024

Abstract

This project aims to predict forest fires using various machine learning models trained on environmental data. By utilizing tools such as Python, Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn, the model assesses factors influencing fire occurrences. Key findings demonstrate the effectiveness of models like Support Vector Classifier, Decision Tree, Gaussian Naive Bayes, and Random Forest in achieving high accuracy.

Introduction

Forest fires pose significant threats to ecosystems and human settlements. Predictive modeling can help mitigate these risks by enabling proactive measures. This project seeks to leverage machine learning to predict the likelihood of forest fires based on historical data and environmental variables.

Objectives:

- To develop a predictive model for forest fire occurrences.
- To evaluate the performance of various classification algorithms.

Importance:

Accurate predictions can assist forest management agencies in resource allocation and risk assessment, ultimately minimizing fire-related damages.

Methodology

The project employs several Python libraries: Pandas for data manipulation, NumPy for numerical operations, Matplotlib and Seaborn for data visualization, and Scikit-learn for machine learning tasks.

Steps Taken:

1. Data Collection: Acquired a dataset containing historical forest fire incidents and relevant environmental factors.

2. Data Preprocessing: Utilized techniques like Train-Test Split, Standard Scaler, and Labelencoder to prepare the data for modeling.

3. Model Training: Implemented various classification models, including:
   - Support Vector Classifier (SVC)
   - Decision Tree Classifier
   - Gaussian Naive Bayes
   - Random Forest Classifier

Implementation

The models were developed using a consistent random state for reproducibility, with a kernel used for the SVC model. The training and testing datasets were carefully split to ensure unbiased performance evaluation.

Challenges:

- Data Imbalance: Addressed by ensuring adequate representation of both fire and non-fire instances.
- Model Selection: Initial attempts with simpler models led to suboptimal results; advanced models were then prioritized.

Results

The performance of each model was assessed using accuracy scores, classification reports, and confusion matrices. Key results include:

- SVC: Achieved an accuracy of 46.92%.
- Decision Tree: Achieved an accuracy of 43%.
- Gaussian Naive Bayes: Achieved an accuracy of 22.29%.
- Random Forest: Achieved the highest accuracy of 50%.

Visualization

Visualizations illustrate model performance comparisons and highlight the importance of specific features in predicting fire occurrences.

Conclusion

The project successfully demonstrates the applicability of machine learning in predicting forest fires. Among the models tested, the Random Forest Classifier provided the most accurate predictions, indicating its potential for practical application in fire management.

Recommendations:

Future work could involve:
- Integrating real-time data for dynamic predictions.
- Exploring more complex models, such as ensemble methods or neural networks.