**A**

**MINI PROJECT REPORT**

**ON**

# “ Forest Fire Confidence Prediction ”

# Data Science and Big Data Analytics

**SUBMITTED BY**

NAME OF STUDENTS (ROLL NO)

2024-25

T.E. (INFORMATION TECHNOLOGY)

Semester - II

Under the Guidance Of

Ms. Rohini Rankhamb



DEPARTMENT OF INFORMATION TECHNOLOGY

MARATHWADA MITRA MANDAL’S COLLEGE OF ENGINEERING, KARVENAGAR, PUNE.

2024-2025

**MARATHWADA MITRA MANDAL’S**

**COLLEGE OF ENGINEERING**

**PUNE.**

**DEPARTMENT OF INFORMATION TECHNOLOGY**



**CERTIFICATE**

This is to certify that student \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, Roll No.\_\_\_\_\_ is studying in Third Year INFORMATION TECHNOLOGY course in SEM II and he/she has successfully completed and submitted the **Mini Project Report under the subject Data Science and Big Data Analytics Laboratory** entitled “ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_” .

This study is a partial fulfillment of the degree of Bachelor of Engineering in Information Technology of Savitribai Phule Pune University, PUNE during the academic year 2024-2025.

Ms. Rohini Rankhamb Dr. Swapnaja Ubale

**Guide**  **H.O.D.**

**ABSTRACT**

Forest fires pose a significant threat to ecological systems, wildlife, and human life. Timely prediction of such events is crucial for effective mitigation and resource planning. This project, titled "Forest Fire Confidence Prediction", focuses on developing a predictive model to estimate the confidence level of forest fire occurrences based on environmental and meteorological data.

Using a dataset containing attributes like temperature, humidity, wind speed, and location coordinates, various data preprocessing and visualization techniques were applied to understand the underlying patterns. Machine learning algorithms such as Logistic Regression, Decision Trees, and Random Forest were implemented to build predictive models. Model performance was evaluated using standard metrics like accuracy, precision, recall, and F1-score.

The results demonstrate the effectiveness of data-driven approaches in predicting fire confidence levels, offering valuable insights that can assist forest management authorities in making informed decisions. This project highlights the practical applications of Data Science and Big Data Analytics in addressing real-world environmental challenges.

**INDEX**

|  |  |  |  |
| --- | --- | --- | --- |
| **CHAPTER NO.** |  | **TOPIC** | **PAGE NO.** |
| **1** |  | **Introduction** | 5 |
|  | 1.1 | Innovation | 6 |
|  | 1.2 | Scope of Project | 7 |
|  | 1.3 | Application of Project | 8 |
| **2** |  | **Project Design** | 9 |
|  | 2.1 | Architecture Diagram | 9 |
| **3** |  | Requirements | 10 |
|  | 3.1 | Software Requirements | 10 |
|  | 3.2 | Hardware Requirements | 10 |
| **4** |  | Sample Source Code | 11 |
| **5** |  | Graphical User Interface | 13 |
| **6** |  | Future Enhancement | 14 |
| **7** |  | Conclusion | 15 |
|  |  | References / Bibliography | 16 |

* 1. **Introduction**
  2. In recent years, the frequency and intensity of forest fires have increased drastically, leading to severe environmental, economic, and social consequences. Forest fires not only destroy vast areas of biodiversity-rich ecosystems but also contribute significantly to carbon emissions and air pollution. Early prediction and assessment of forest fire confidence levels can play a vital role in forest management, disaster response, and minimizing potential damages.
  3. The objective of this project is to build a predictive model that can estimate the confidence level of a forest fire based on various environmental and meteorological parameters. By leveraging data science and big data analytics techniques, the project aims to analyze large datasets, identify key features influencing fire occurrences, and apply suitable machine learning algorithms to predict the likelihood or confidence of a fire event.
  4. The dataset used in this project contains real-world records of forest fire incidents with attributes such as temperature, humidity, wind speed, and spatial coordinates. Through exploratory data analysis, preprocessing, model training, and evaluation, the project provides insights into the conditions leading to high-confidence fire events and helps in developing a decision-support system for authorities and forest departments.
  5. This project not only demonstrates the practical application of data science in solving real-world environmental challenges but also highlights the importance of predictive analytics in proactive disaster management.
  6. **1.1 Innovation :**

The Forest Fire Confidence Prediction project includes several innovative approaches that enhance the accuracy, efficiency, and practicality of the prediction system. These innovations are:

1. Intelligent Feature Selection

The project focuses on selecting the most relevant features such as temperature, humidity, wind speed, and location data to improve the performance of the predictive model.

2. Confidence Level-Based Prediction

Instead of simply predicting whether a fire will occur or not, the model predicts the confidence level of a forest fire, which offers more detailed and actionable insights.

3. Model Comparison and Evaluation

Multiple machine learning algorithms like Logistic Regression, Decision Tree, and Random Forest were implemented and compared to find the best-performing model for this task.

4. Risk Classification System

Based on the predicted confidence values, the model categorizes fire risk into levels such as low, medium, and high. This helps in prioritizing preventive actions.

5. Modular and Scalable Design

The system is designed in a modular way, making it easier to scale or integrate with real-time data sources like weather APIs or sensor data in the future.

**1.2 Scope of Project :**

The project "Forest Fire Confidence Prediction" aims to leverage data science and machine learning techniques to build a predictive system that can estimate the likelihood or confidence of forest fire occurrences based on key environmental factors. The scope of this project is outlined as follows:

* **Data Collection and Preprocessing**: Utilize a real-world dataset containing forest fire records along with relevant meteorological attributes such as temperature, humidity, wind speed, and geographical coordinates.
* **Exploratory Data Analysis (EDA)**: Perform statistical analysis and data visualization to identify patterns, trends, and correlations between features and forest fire confidence levels.
* **Model Building**: Apply and compare different machine learning algorithms (e.g., Logistic Regression, Decision Tree, Random Forest) to develop a reliable prediction model.
* **Performance Evaluation**: Evaluate the predictive models using metrics like accuracy, precision, recall, F1-score, and confusion matrix to determine their effectiveness.
* **Practical Application**: Provide insights that can help forest management departments and disaster response teams in assessing fire risks and implementing timely preventive measures.
* **Scalability**: Design the solution in a way that it can be scaled or enhanced in the future by incorporating additional features such as satellite imagery, vegetation index, or real-time sensor data.

**1.3 Application of Project :**

The "Forest Fire Confidence Prediction" project has several real-world applications, particularly in the domains of environmental monitoring, disaster management, and public safety.

Key applications include:

1. **Early Warning Systems**  
   Predicting the confidence level of forest fires allows authorities to issue timely alerts and warnings, reducing damage to ecosystems and loss of life.
2. **Forest Resource Management**  
   Helps forest departments in monitoring high-risk areas and planning fire-prevention activities like controlled burning or deploying firewatch personnel.
3. **Disaster Preparedness and Response**  
   Enables better allocation of emergency response teams and firefighting equipment to regions with high fire risk.
4. **Environmental Conservation**  
   Assists in protecting endangered flora and fauna by forecasting potential threats and allowing conservationists to act proactively.
5. **Smart Surveillance and Monitoring**  
   Can be integrated with IoT-based sensor systems or drones to provide real-time prediction and automated decision-making.
   1. **2.Project Design**
   2. **2.1 Architecture Diagram :**

**3.Requirements**

**3.1 Software Requirements :**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Component** | **Minimum Requirement** |
| 1 | Processor | Intel Core i3 or equivalent |
| 2 | RAM | 4 GB (8 GB recommended for smoother development) |
| 3 | Hard Disk | 2 GB of free space |
| 4 | Display | 1024×768 resolution or higher |

**3.2 Hardware Requirements :**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Software/Tool** | **Description** |
| 1 | Operating System | Windows 10/11, macOS, or Linux |
| 2 | Code Editor | VS Code / Sublime Text / Atom |
| 3 | Browser | Chrome / Firefox / Edge (for testing) |
| 4 | Version Control | Git |

**4.Sample Source Code**

**A screenshot of a computer

AI-generated content may be incorrect.**

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**5.Graphical User Interface**

**A screenshot of a computer

AI-generated content may be incorrect.**

A screenshot of a computer

AI-generated content may be incorrect.

**6. Future Enhancement**

**Future Enhancements**

While the current model for Forest Fire Confidence Prediction provides valuable insights and predictive capabilities, there are several areas where the project can be further enhanced for improved accuracy, scalability, and real-world application. Some possible future enhancements include:

**1.**Integration with Real-Time Data Sources:

Incorporating real-time meteorological data, satellite feeds, and IoT-based environmental sensors can significantly improve the responsiveness and accuracy of predictions.

**2.**Use of Advanced Machine Learning Models:

Implementing more sophisticated algorithms such as Gradient Boosting, XGBoost, or deep learning models like Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) for image-based inputs.

**3.**Geospatial Analysis and Visualization:

Enhancing the project with GIS (Geographic Information System) tools to visualize high-risk zones on maps and perform spatial risk assessments.

**4.**Mobile and Web Application Development:

Creating a user-friendly interface for forest officials and the public to access predictions, alerts, and visualizations easily.

**5.**Multilingual Alert System:

Developing an alert system that can send SMS or push notifications in multiple languages to notify local communities about potential fire risks.

**7. Conclusion**

The Forest Fire Confidence Prediction project successfully demonstrates how data science and machine learning techniques can be applied to address real-world environmental challenges. By analyzing historical forest fire data and environmental parameters, the project provides a predictive model capable of estimating the confidence level of fire occurrences. This can play a crucial role in early detection, proactive prevention, and effective disaster management.

Throughout the project, essential steps such as data preprocessing, exploratory data analysis, model building, and performance evaluation were carried out systematically. Various machine learning algorithms were implemented and compared to identify the most accurate and reliable model.

The findings of this project offer valuable insights for forest management authorities, enabling them to make informed decisions and take timely preventive measures. While the current model lays a strong foundation, future enhancements—such as real-time data integration and geospatial visualization—can significantly improve its effectiveness and practical applicability.

Overall, this mini project highlights the powerful role of data science in solving critical environmental issues and contributes meaningfully to the field of predictive analytics.

**8. References**

**1.** Cortez, P., & Morais, A. (2007). A Data Mining Approach to Predict Forest Firesusing Meteorological Data. Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence.

**2.** Scikit-learn: Machine Learning in Python. <https://scikit-learn.org>

**3**. Pandas Documentation – Data Analysis Library in Python. https://pandas.pydata.org

**4**. Matplotlib Documentation – Python Plotting Library. <https://matplotlib.org>

**5**. Seaborn: Statistical Data Visualization in Python. https://seaborn.pydata.org

**6.** UCI Machine Learning Repository – Forest Fires Data Set. <https://archive.ics.uci.edu/ml/datasets/Forest+Fires>

**7**. Jupyter Notebook – Project Development Environment. <https://jupyter.org>

**8.** Kaggle – Forest Fire Data and ML Model Examples. <https://www.kaggle.com>