Tempest FWI Predictor

Project Overview

Project Name: Tempest FWI Predictor – A Machine Learning Model to Predict Fire Weather Index

Objective: Develop a machine learning model that accurately predicts the Fire Weather Index (FWI) using environmental and meteorological data. This model aims to provide early warning capabilities for fire risk assessment based on weather conditions.

Status: Data Preprocessing Milestone (In Progress)

Project Scope

This project focuses on building a predictive model for forest fire risk assessment. The Fire Weather Index (FWI) is a critical metric used to estimate fire behavior based on atmospheric conditions and fuel characteristics.

Key Features

- **Target Variable:** Fire Weather Index (FWI)
- **Data Source:** Forest fires dataset
- **Primary Variables:** Temperature, Relative Humidity, Wind Speed, Rainfall, FFMC (Fine Fuel Moisture Code), DMC (Duff Moisture Code), ISI (Initial Spread Index)
- **Geographic Region:** Data includes regional/location-based features

---

Current Milestone: Data Preprocessing

Milestone Objectives

The preprocessing phase aims to prepare raw data for analysis and modeling through the following steps:

1. Data Collection

- **Status:** Complete
- **Action:** Downloading and accessing FWI and environmental data
- **Output:** `forestfires.csv`
- **Description:** Raw dataset containing meteorological and FWI variables collected from forest fire records

2. Data Cleaning

- **Status:** In Progress
- **Actions:**
  - Handle missing values (NaN) by replacing with 0 or mean values
  - Convert categorical variables (month, day) into numerical codes
  - Validate data integrity and consistency
- **Output:** Cleaned dataset ready for analysis
- **Categorical Variables:** Month (encoded as numeric), Day of week (encoded as numeric)

3. Data Analysis

- **Status:** In Progress
- **Actions:**
  - Generate histograms for all numeric features
  - Calculate correlation matrix and covariance
  - Create correlation heatmap for feature relationship visualization
  - Identify feature distributions and potential outliers
- **Output:** `forestfires_processed.csv`, Statistical summaries, Visualization plots

---

Technical Implementation

Technology Stack

Component	Technology
Language	Python 3.x
Data Processing	Pandas, NumPy
Visualization	Matplotlib, Seaborn
Environment	Miniconda/Conda virtual environment
IDE	VS Code
Version Control	GitHub (project repository to be shared)

Data Processing Pipeline

Raw Data (forestfires.csv) ↓ Load with Pandas ↓ Handle Missing Values (fillna with mean/0) ↓ Encode Categorical Variables (month, day → numeric codes) ↓ Data Validation (check for remaining NaN values) ↓ Exploratory Data Analysis (histograms, correlation) ↓ Export Processed Data (forestfires_processed.csv)

Key Code Functions

Data Loading & Cleaning:

```python
# Load dataset
df = pd.read_csv('forestfires.csv')

# Encode categorical variables
df['month'] = df['month'].astype('category').cat.codes
df['day'] = df['day'].astype('category').cat.codes

# Handle missing values
for col in df.columns:
    if df[col].dtype in ['float64', 'int64']:
        df[col].fillna(df[col].mean(), inplace=True)
```

Data Analysis:

```python
# Correlation matrix
corr = df.corr()
print(corr)

# Visualization
df.hist(bins=15, figsize=(15, 12))
sns.heatmap(corr, annot=True, fmt=".2f", cmap="coolwarm")

# Export processed data
df.to_csv("forestfires_processed.csv", index=False)
```

Features & Variables

Input Features

Feature	Type	Description
Temperature	Numeric	Atmospheric temperature in degrees
Relative Humidity	Numeric	Moisture content in air (0-100%)
Wind Speed	Numeric	Speed of wind at time of measurement
Rainfall	Numeric	Amount of rain recorded
FFMC	Numeric	Fine Fuel Moisture Code (fuel dryness indicator)
DMC	Numeric	Duff Moisture Code (intermediate layer moisture)
ISI	Numeric	Initial Spread Index (fire spread potential)
Month	Categorical → Numeric	Month of occurrence (encoded as numeric)

Feature	Type	Description
Day	Categorical → Numeric	Day of week (encoded as numeric)
Region	Categorical	Geographic region of fire occurrence

Target Variable

Variable	Description
FWI (Fire Weather Index)	Numerical index representing fire risk level; combination of FFMC, DMC, and ISI with fire behavior calculations

---

Project Learning Objectives

Week 1-2 Learning Outcomes

Students should gain understanding in the following foundational concepts:

1. Linear Regression Fundamentals

- Understand the theory and assumptions behind linear regression
- Learn how linear relationships between variables are modeled
- Application to FWI prediction as a regression problem

2. Solution Methods for Regressors

- Ordinary Least Squares (OLS) method for parameter estimation
- Gradient descent optimization approaches
- Maximum likelihood estimation principles
- Regularization techniques (Ridge, Lasso)

3. Systems of Linear Equations

- Matrix algebra fundamentals
- Solving systems of linear equations
- Eigenvalues and eigenvectors
- Application in machine learning model development

Project Structure

Repository Organization

```
Tempest-FWI-Predictor/
  README.md                          # Project overview
  data/
     forestfires.csv                 # Raw dataset
     forestfires_processed.csv       # Cleaned/processed dataset
  notebooks/
     [Your-Name]/
         data_preprocessing.py       # Data cleaning pipeline
         data_analysis.py            # EDA and visualization
  src/
     preprocessing.py                # Reusable preprocessing functions
     analysis.py                     # Analysis utilities
     config.py                       # Configuration parameters
  results/
     correlation_matrix.csv          # Calculated correlations
     visualizations/                 # Generated plots
  docs/
     PROJECT_DOCUMENTATION.md        # This file
  .gitignore                         # Git ignore rules
```

Important Notes

- **Individual Folders:** Each team member should create a separate folder under the repository with their name
- **File Organization:** Include all preprocessing and analysis files in your designated folder
- **GitHub Repository:** The project repository link will be shared by the project lead
- **Collaboration:** Follow the established file structure for team consistency

---

Current Status & Deliverables

Completed Tasks

- Project initialization and planning
- Dataset identification (forestfires.csv)
- Development environment setup (Python, Pandas, NumPy, Matplotlib, Seaborn)
- Data loading implementation
- Initial data exploration

In Progress

- Data cleaning and missing value handling
- Categorical variable encoding
- Exploratory Data Analysis (EDA)
- Correlation and distribution analysis

Upcoming Milestones

- Feature engineering and selection
- Linear regression model development
- Model training and validation
- Performance evaluation and optimization
- Hyperparameter tuning
- Final model deployment and documentation

Installation & Setup

Prerequisites

- Python 3.7 or higher
- Conda/Miniconda package manager
- Git for version control

Environment Setup

```
# Create conda environment
conda create -n tempest-fwi python=3.9

# Activate environment
conda activate tempest-fwi

# Install required packages
pip install pandas numpy matplotlib seaborn scikit-learn jupyter

# Or use conda
conda install -c conda-forge pandas numpy matplotlib seaborn scikit-learn
```

Running the Code

```
# Navigate to project directory
cd Tempest-FWI-Predictor

# Run preprocessing script
python notebooks/[Your-Name]/data_preprocessing.py
```

```
# Run analysis script
python notebooks/[Your-Name]/data_analysis.py
```

Expected Outcomes

Phase 1: Preprocessing & Analysis (Current)

Deliverables: - Cleaned dataset without missing values or encoding issues - Statistical summaries (mean, median, std deviation, quartiles) - Feature distribution histograms (visual exploration) - Correlation matrix showing feature relationships - Correlation heatmap visualization - Processed CSV file for downstream analysis

Quality Metrics: - Zero missing values in final dataset - All categorical variables successfully encoded - Clear identification of feature importance through correlation - No data loss during preprocessing

---

References & Resources

Documentation

- Pandas Documentation: https://pandas.pydata.org/docs/
- NumPy Documentation: https://numpy.org/doc/
- Seaborn Visualization: https://seaborn.pydata.org/
- Scikit-learn ML Library: https://scikit-learn.org/

Theory Resources

- Linear Regression Basics
- Fire Weather Index (FWI) System Overview
- Machine Learning Preprocessing Best Practices
- Statistical Analysis Methods

Contact & Collaboration

- **Project Lead:** [To be provided]
- **Repository:** [GitHub link to be shared]
- **Communication:** [Team communication channel]
- **Submission Date:** [As per project timeline]

Revision History

Version	Date	Author	Changes
1.0	2025-11-21	Project Documentation	Initial documentation creation - Project scope and objectives - Milestone definitions - Technical implementation details - Feature descriptions

---

Last Updated: November 21, 2025
Status: Active Development
Next Review: Upon completion of data preprocessing milestone