Weather Trend Forecasting Project Report

By: Panchal Siddh

1. Introduction

Objective:

The goal of this project is to analyze global weather trends, detect anomalies, and improve forecasting accuracy using machine learning models. This study helps in understanding climate patterns and their environmental impact.

Key Goals:

- Perform data cleaning and handle missing values, outliers, and normalization.
- Conduct Exploratory Data Analysis (EDA) to uncover trends and correlations.
- Build and compare multiple forecasting models (ARIMA, Exponential Smoothing, Linear Regression, Ensemble).
- Implement anomaly detection for weather pattern irregularities.
- Conduct climate analysis, spatial analysis, and environmental impact studies.

2. Dataset Overview

- Source: Kaggle – Global Weather Repository
- Features: Over 40 attributes including temperature, humidity, precipitation, wind speed, air quality, etc.
- Data Size: Large dataset covering global weather records.

3. Data Cleaning & Preprocessing

- Handled missing values using median/mode imputation.
- Normalized numerical features with StandardScaler.
- Removed outliers using the IQR method.
- Converted timestamps into a structured datetime format for time-series analysis.

4. Exploratory Data Analysis (EDA)

- Temperature Trends: Long-term patterns visualization.
- Precipitation Analysis: Correlation with temperature & humidity.
- Heatmap Analysis: Identifies relationships between weather parameters.

5. Anomaly Detection

- Algorithm Used: Isolation Forest
- Key Findings:
    - Identified extreme temperature fluctuations.
    - Detected unexpected precipitation patterns.
    - Flagged anomalies in air quality data.

6. Forecasting Models

- ARIMA: Predicts future temperature values based on time-series analysis.
- Exponential Smoothing: Captures seasonality and trends.
- Linear Regression: Establishes a baseline forecast.
- Ensemble Model: Combines predictions from all models for improved accuracy.

Performance Metrics:

Model	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
ARIMA	0.8766	1.141
Exponential Smoothing	0.7954	0.9728
Linear Regression	0.7902	0.9679
Ensemble Model (Best Accuracy)	0.8091	0.9936

7. Climate Analysis & Environmental Impact

- Climate Change Trends: Rising temperatures observed over decades.
- Air Quality Correlation: Strong relationship between pollution levels and weather conditions.
- Spatial Analysis: Visualized temperature variations across different regions.

8. Conclusion & Future Work

Conclusion:

The Ensemble Model provided the most accurate forecasting results. Anomaly detection successfully identified extreme weather patterns. Climate analysis revealed significant global temperature variations.

Future Improvements:

- Implement Deep Learning Models (LSTMs, Transformer-based models).
- Hyperparameter tuning for improved forecasting accuracy.
- Real-time API integration for live weather predictions.

9. References & Acknowledgements

- Dataset: Kaggle Global Weather Repository
- Libraries Used: Pandas, NumPy, SciPy, TensorFlow, GeoPandas