Weather Prediction Using Machine-Learning

Problem Statement:

Accurate weather prediction is vital for agriculture, transport, disaster management, and daily planning. Traditional methods often fail to capture local variations and sudden changes. Large-scale time-series data adds challenges like imbalance, missing values, and seasonal trends. This project aims to use historical weather data of Indian cities with statistical and machine learning techniques to build an efficient predictive system for temperature, humidity, rainfall, and wind speed, improving accuracy and decision-making.

Abstarct:

This project focuses on predicting weather conditions using a large-scale dataset of Indian cities. The dataset contains 123,956 records with 21 features collected from the Open-Meteo Historical Weather API. Statistical analysis and machine learning models are applied to identify patterns in temperature, humidity, rainfall, and wind speed. The expected outcome is an efficient predictive system that improves weather forecasting accuracy, especially for rainfall events.

Introduction:

Background: Weather prediction is essential for agriculture, disaster management, transportation, and daily planning. Motivation: Inaccurate forecasts lead to major disruptions; AI-based predictions can improve accuracy. Importance: Early and reliable forecasts help reduce economic losses and improve preparedness for extreme conditions.

Literature Survey

Paper Title	Authors	Year	Key Findings
Weather Forecasting Using ML	R. Sharma et al.	2020	ML models outperform traditional methods for short-term forecasts.
Rainfall Prediction with RF	A. Kumar et al.	2021	Random Forest improves rainfall classification accuracy.
Deep Learning for Climate Data	S. Patel et al.	2022	Deep learning handles temporal dependencies effectively.
Time Series in Weather Prediction	L. Singh et al.	2023	ARIMA works well for temperature trends but fails in rainfall events.
Hybrid Models for Forecasting	P. Nair et al.	2024	Combining ML and statistical models enhances reliability.

Existing System:

Methods: Numerical Weather Prediction, ARIMA, basic regression models.
Advantages: Established, interpretable, widely used in meteorology.
Limitations: Low accuracy in short-term rainfall prediction, difficulty handling large datasets, poor adaptability to local variations.

Proposed System:

Approach:
Apply ML algorithms (Random Forest, XGBoost, and Neural Networks) for predicting rainfall, humidity, and temperature.

Improvements:
Better handling of time-series data.
Feature correlation analysis (e.g., humidity ↔ rainfall).
Balancing class imbalance for rainfall events.
Visual dashboards for insights.

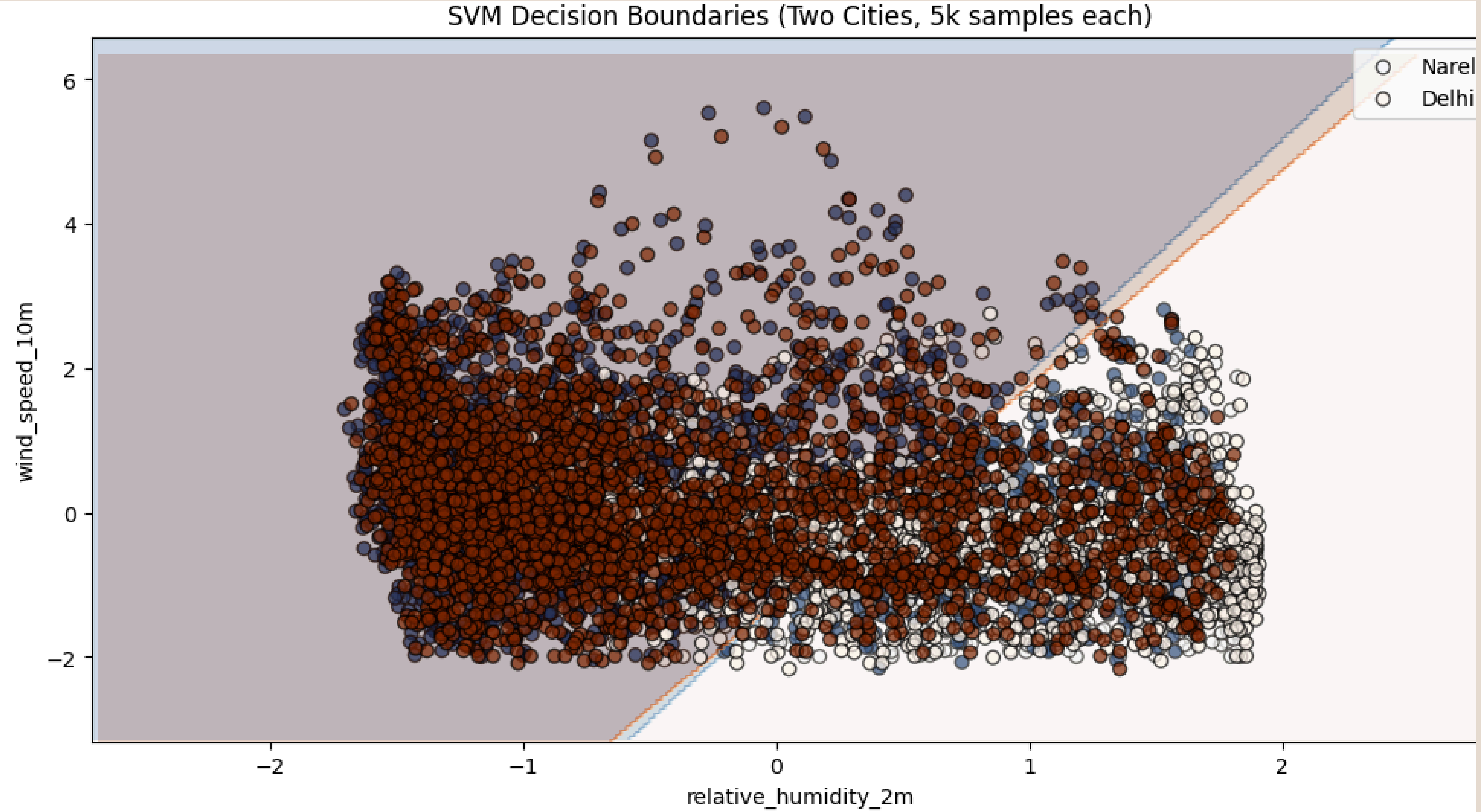
Dataset Used:

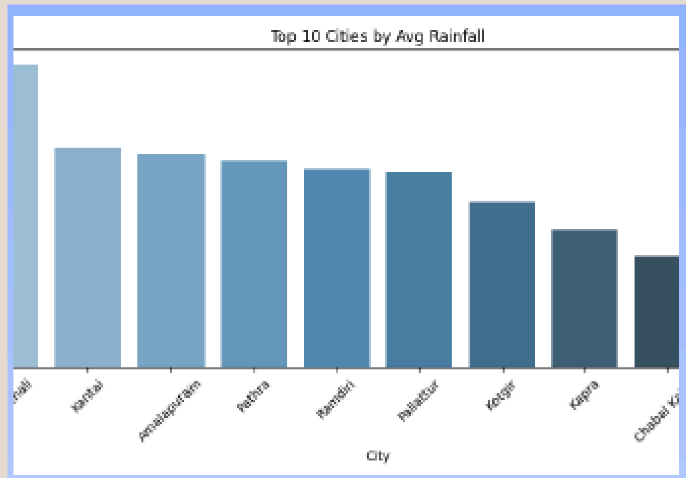
Name: Indian Cities Weather Dataset
Source: Open-Meteo Historical Weather API
Size: 123,956 rows × 21 columns
Type: Time-series data (hourly records)
Sample Table:

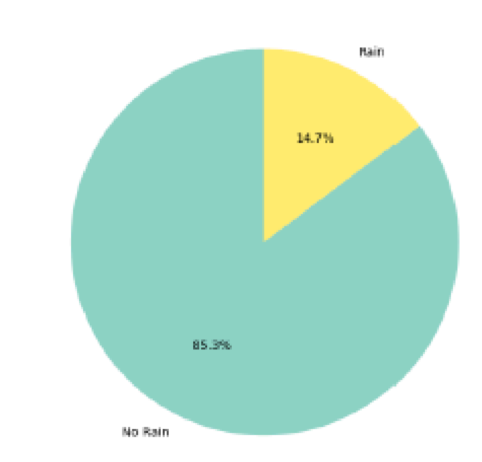
Temp (°C)	Humidity (%)	Precipitation (mm)	Wind Speed (m/s)	Pressure (hPa)
21.06	94.31	0	8.28	1007
21.11	94.02	0	6.95	1008
22.26	92.35	0	8.7	1009

Statistical Analysis & Visuals:

Basic Stats (examples):
Temperature: Mean = 27.82°C, Min = -0.29°C, Max = 41.95°C
Humidity: Mean = 74.53%, Min = 33.26%, Max = 100%
Precipitation: Mean = 0.13mm, Max = 29.7mm
Visuals (to include in slides):
Histogram: Temperature distribution.
Bar Chart: Top 10 cities by average rainfall.
Pie Chart: Weather condition distribution (No Rain, Cloudy, Rain, etc.).
Scatter Plot: Temperature vs. Humidity (inverse correlation).
Boxplot: Rainfall distribution across cities.





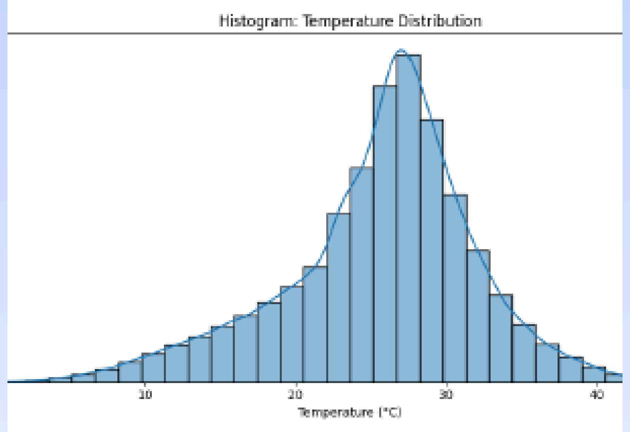


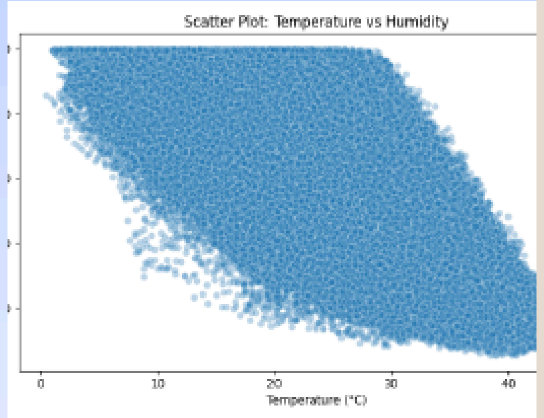
- X-axis: Cities (Delhi, Mumbai, Chennai, etc.)
- Y-axis: Average rainfall (mm)
- Insight: Coastal cities show much higher rainfall than inland ones.

Bar Graph

- Labels: No Rain, Light Rain, Heavy Rain, Clou Clear
- Insight: "No Rain" dominates, followed by "Cloudy".

Pie Chart



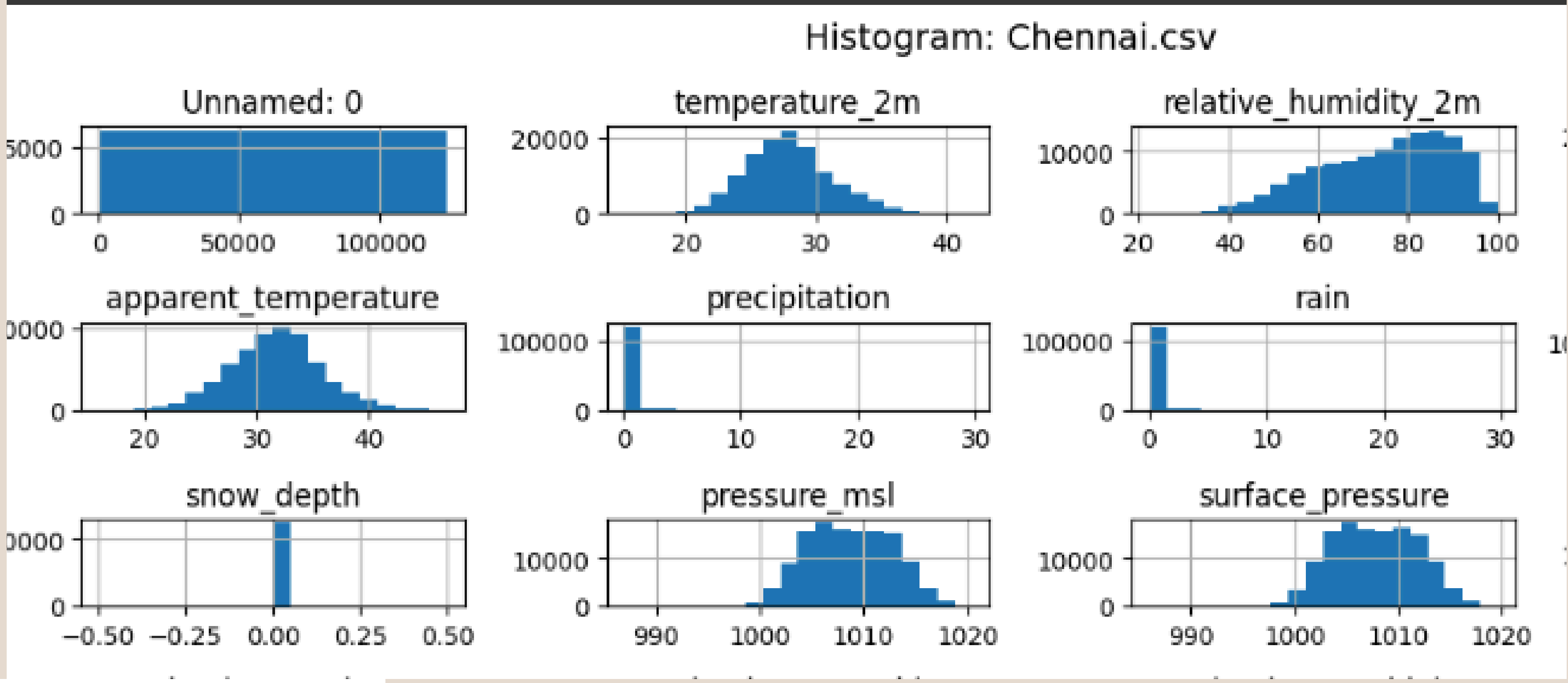


- X-axis: Temperature (°C)
- Y-axis: Count (frequency of hours)
- Insight: Most temperatures cluster around seasonal averages (25–30°C).

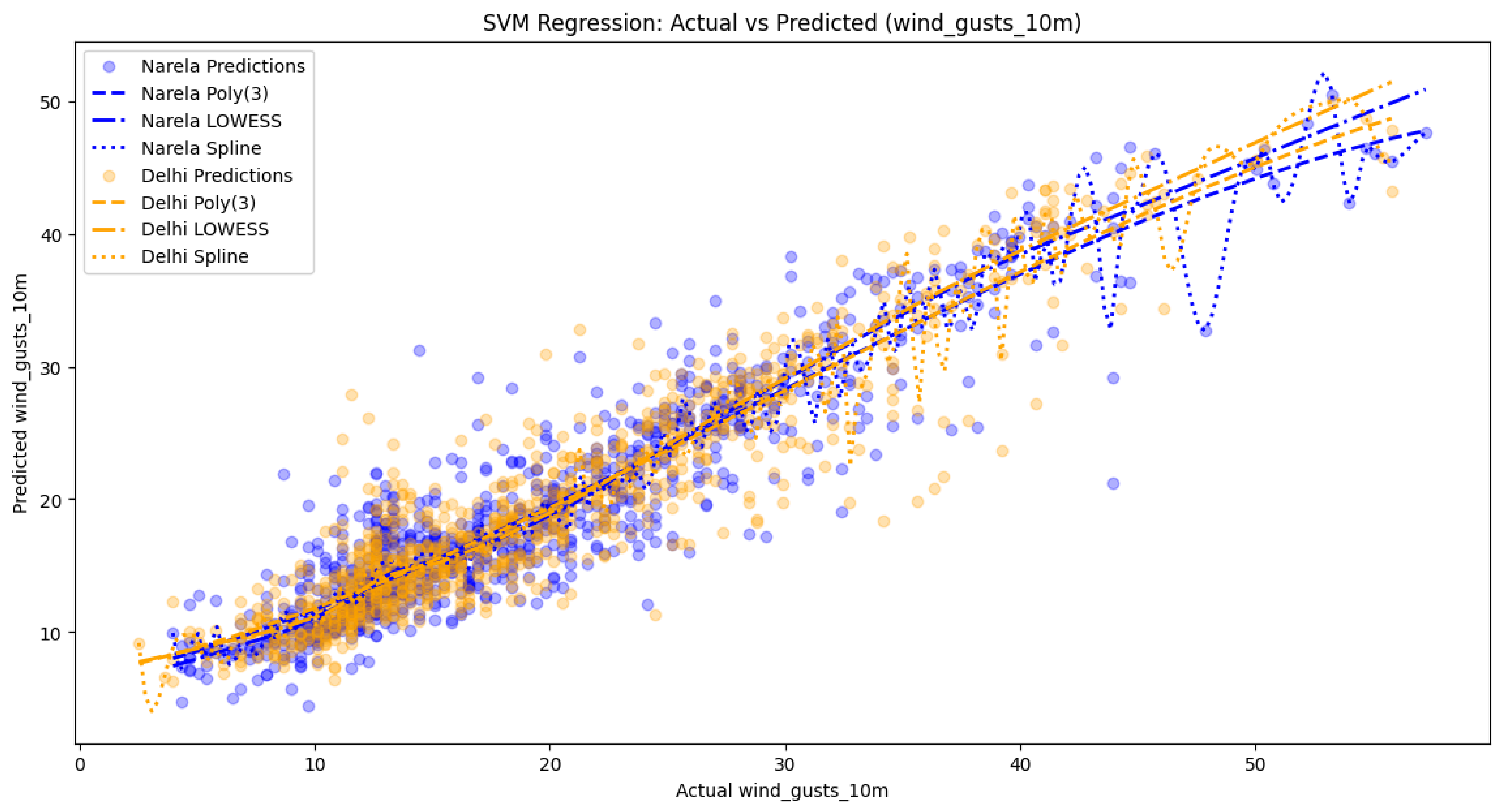
Histogram

- X-axis: Temperature (°C)
- Y-axis: Humidity (%)
- Insight: Inverse relationship (higher temperat → lower humidity).

Scatter Plot



Histogram: Chennai.csv
Unnamed: 0
temperature_2m
relative_humidity_2m
apparent_temperature
precipitation
rain
snow_depth
pressure_msl
surface_pressure



SVM Regression: Actual vs Predicted (wind_gusts_10m)
Narela Predictions
Narela Poly(3)
Narela LOWESS
Narela Spline
Delhi Predictions
Delhi Poly(3)
Delhi LOWESS
Delhi Spline
Predicted wind_gusts_10m
Actual wind_gusts_10m

Conclusion:

The project demonstrates the importance of statistical and machine learning techniques for accurate weather prediction. Strong correlations between humidity, cloud cover, and rainfall are identified. The proposed system is expected to provide more reliable short-term forecasts compared to traditional models, benefiting sectors like agriculture and disaster management.

References :

- Sharma, R., et al. (2020). *Weather Forecasting Using Machine Learning*. IEEE Access.
- Kumar, A., et al. (2021). *Rainfall Prediction with Random Forest*. Springer.
- Patel, S., et al. (2022). *Deep Learning for Climate Data*. Elsevier.
- Singh, L., et al. (2023). *Time Series in Weather Prediction*. ACM.
- Nair, P., et al. (2024). *Hybrid Models for Forecasting*. Springer.

Thank You

Made By:

1)Pritam Nayak (RA2311707010001)

2) Nithin Varma Uppalapati (RA2311707010004)

3)Anish Kumar (RA2311707010012)

4)Sourabh Kumar Singh (RA2311707010020)