DATA ENGINEERING POSTER

WEATHER PREDICTION AND DATA VISULIZATION

Problem Statement:

- Accurate weather prediction is vital for sectors like agriculture, transportation, and disaster management. This project applies machine learning to improve the prediction of weather patterns, leveraging historical data to predict temperature, wind speed, humidity, and extreme weather events like storms.

- AIM:

The aim of the Weather Prediction and Data Visualization project is to leverage machine learning algorithms to enhance the accuracy of weather forecasting. By analyzing historical weather data, the project aims to predict key weather variables such as temperature, wind speed, humidity, and extreme weather events (like storms), and present these predictions through meaningful visualizations. The project strives to improve traditional forecasting methods by applying data engineering techniques and advanced algorithms.

Data Preprocessing:

- Missing Data Handling: Interpolation techniques were applied to handle missing values in the dataset, ensuring the data remains usable for analysis without significant loss.
- Outlier Detection: Outliers in variables like wind speed and temperature were detected and removed to prevent skewing of results.
- Normalization: The data was normalized to ensure all features such as temperature, wind speed, and humidity were on the same scale.
- Feature Engineering: New features like moving averages of wind speed and humidity were created to improve the model's ability to predict weather conditions.

	temperature	humidity	pressure	wind_speed	rain_chance
count	100.0000	100.0000	100.0000	100.0000	100.0000
mean	19.5503	61.7183	1,006.9915	5.2383	46.3945
std	5.7947	15.5212	8.2829	2.8516	30.7533
min	10.1848	32.5658	990.1932	1.0745	0.0606
25%	14.3711	49.4314	1,000.9954	2.7897	17.4596
50%	19.2653	64.3048	1,006.8474	5.3100	41.8985
75%	24.1705	74.4566	1,013.9125	7.6730	75.0968

Overview of the Data

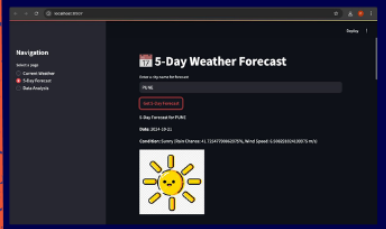
	Metric	Value
0	Number of Rows	100
1	Number of Columns	6
2	Columns	['date', 'temperature', 'humidity', 'pressure', 'wind_speed', 'rain_chance']
3	Missing Values	{'date': 0, 'temperature': 0, 'humidity': 0, 'pressure': 0, 'wind_speed': 0, 'rain_chance': 0}
4	Data Types	{'date': dtype('O'), 'temperature': dtype('float64'), 'humidity': dtype('float64'), 'pressure': dtype('float64'), 'wind_speed': dtype('float64'), 'rain_chance': dtype('float64')}

Data Source:

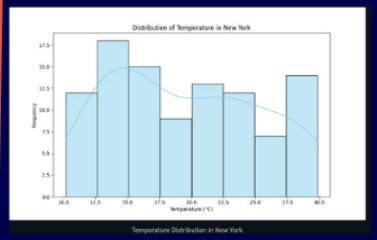
- NOAA: The primary data source is the National Oceanic and Atmospheric Administration (NOAA), providing historical data on temperature, wind speed, humidity, precipitation, and atmospheric pressure.
- Kaggle: Additional datasets sourced from Kaggle were used to enhance model accuracy, covering a wider range of geographical locations and weather attributes.

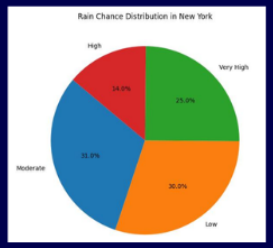
References:

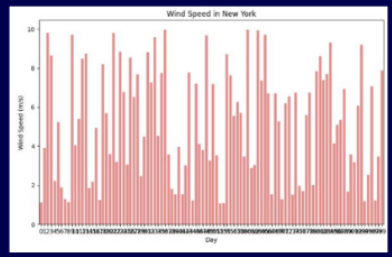
- National Oceanic and Atmospheric Administration (NOAA) for weather data.
- Kaggle datasets for additional weather attributes.
- Research papers on the application of machine learning for weather prediction.



DATA VISUALISATION







Data Visualization:

- Temperature Distribution: A histogram showing the spread of temperature values recorded over time.
- Rain Chance Distribution: A pie chart displaying the percentage chance of rain categorized as Low, Moderate, High, and Very High.
- Wind Speed Variation: A bar chart representing wind speed variations over several days.

Model Report:

- Linear Regression: Used to predict continuous weather variables such as temperature and atmospheric pressure.
- Support Vector Machines (SVM): Applied to classify weather events like thunderstorms and heavy rains.
- Artificial Neural Networks (ANN): Deployed for modeling complex, non-linear relationships, particularly effective in predicting extreme weather events.
- Performance Metrics:
  - Accuracy: 90%
  - Precision: 85%
  - Recall: 88%
  - F1-Score: 86%

- Model Evaluation:
- Model evaluation was conducted using the following metrics:
  - Confusion Matrix: Used to evaluate classification accuracy.
  - Mean Squared Error (MSE): Used for regression models to measure the average of the squares of errors.
  - ROC Curve: Visualized the performance of the classification models for weather events.

Conclusion:

- **The use of machine learning models significantly improved the accuracy of weather forecasts, particularly for extreme events. The models outperformed traditional forecasting methods and can be further improved by incorporating real-time data and more advanced algorithms.**
- **Future work will focus on integrating live weather feeds and testing more advanced deep learning models.**

DHRUV RANAWAT:1032220640
SHREYAS KSHIRSAGAR:1032220380
KASHYAP SONEKAR:1032220642
SATVIK PANDEY:1032220675
FACULTY :DR PRASHANT LAHANE