**Report On Big Data Analysis Project**

This project demonstrates the effective use of LSTM for predicting time-series data, with clear visualizations to interpret model accuracy and performance in wind speed forecasting.

**1. Abstract**

This project focuses on predicting wind speeds using a **Long Short-Term Memory (LSTM)** neural network. The model uses historical wind speed data to train and evaluate its performance. This approach highlights the use of deep learning for time-series data, leveraging the sequential nature of wind speed measurements to predict future values. The project utilizes data pre-processing techniques, model training, and evaluation, presenting an accessible yet effective model for real-world wind speed forecasting applications.

**2. Explanation of Output**

The project produces three main outputs:

1. **Historical Wind Speed Plot**: Displays the original wind speed data from the dataset, helping to identify trends or anomalies.
2. **Scatter Plot of Predictions**: Compares actual versus predicted wind speed values. Ideally, points close to a 45-degree line indicate a high degree of accuracy in the model's predictions.
3. **Line Plot of Real vs. Predicted Data**: Provides a time-series view comparing the real wind speed values and the model’s predicted values, helping assess the model's ability to capture patterns in the data.

**3. Algorithm Used: LSTM (Long Short-Term Memory)**

The LSTM algorithm is well-suited for time-series prediction tasks due to its ability to learn from past data while maintaining information across longer sequences. Key steps in the project include:

* **Data Preparation**:
  + Extracts the target variable and creates feature arrays from historical wind speed data.
  + Scales features using **MinMaxScaler** to improve training efficiency.
* **LSTM Model Structure**:
  + **LSTM Layer**: Has 10 units with the tanh activation function and hard\_sigmoid for recurrent activations.
  + **Dense Layer**: Outputs the final prediction as a single value (predicted wind speed).
* **Model Compilation**:
  + Uses **Mean Squared Error (MSE)** as the loss function and **RMSprop** as the optimizer.
* **Model Training**:
  + Trained for 25 epochs with verbose output to track MSE at each epoch.

**4. Explanation of Output Plots**

1. **Wind Speed Plot**: Shows the original wind speed data over time, helping visualize patterns that the model will learn. Peaks and valleys reveal fluctuations in speed, which the model attempts to predict.
2. **Scatter Plot of Predictions**: This plot compares actual values (Y\_test) against predictions (predict). If the model is accurate, data points cluster around the diagonal, showing close alignment between actual and predicted values.
3. **Line Plot of Real vs. Predicted Data**: This plot overlays the actual and predicted wind speeds, allowing a side-by-side comparison over time. A good model fit is indicated by the predicted line closely following the actual data line, showing the model's capacity to capture trends and variances in wind speed.