## 

Deep Learning for Weather Prediction

Bharath Reddy Jakkidi

A Project Submitted to

GRAND VALLEY STATE UNIVERSITY

In

Partial Fulfillment of the Requirements

For the Degree of

Master of Science in Applied Computer Science

School of Computing

April 2025



The signatures of the individuals below indicate that they have read and approved the project of Bharath Reddy Jakkidi in partial fulfillment of the requirements for the degree of Master of Science in Applied Computer Science.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Yong Zhuang, Project Advisor Date

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Robert Adams, Graduate Program Director Date

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 Dr. Marouane Kessentini, Unit head Date

# Abstract

This project presents a weather forecasting system that uses recent data and machine learning to predict daily weather conditions. It focuses on analyzing the last seven days of weather data including temperature, humidity, rainfall, sunshine duration, and wind patterns to make accurate predictions for the next day. The model used is an LSTM (Long Short-Term Memory), a type of neural network that is well-suited for time series data. It also considers geographical factors like latitude, longitude province, region, and station, as well as date-related features such as the day of the week, to improve prediction accuracy. The system predicts key weather elements such as minimum and maximum temperature, average humidity, sunshine duration, rainfall, wind speed, and direction. This tool can be especially useful for farmers, delivery services, and anyone who relies on weather forecasts for daily planning. By automating the process and providing consistent, short-term forecasts, the system supports informed decision-making in various weather-dependent activities.

# ****Introduction****

The project at hand addresses a critical challenge in weather forecasting: the prediction of various weather parameters such as temperature, humidity, rainfall, wind speed, and wind direction based on historical weather data. The ability to predict weather conditions with high accuracy can significantly impact industries such as agriculture, delivery services, and event planning, where weather is a key determinant of operations.

The project aims to build a predictive model using machine learning techniques to generate reliable forecasts of weather conditions for future dates, enabling stakeholders to make informed decisions and mitigate risks associated with unpredictable weather.

# ****Project Justification and Objectives****

## ****Research Problem and Motivation****

The research problem stems from the need to develop a more automated and accessible method for weather forecasting that leverages historical data. Traditional weather prediction models can be computationally expensive and require significant human intervention. The goal of this project is to automate the forecasting process using a dataset that includes weather conditions like temperature, humidity, wind speed, and other key metrics.

The motivation behind the project lies in improving the precision and accessibility of weather data for everyday use, especially for industries that heavily rely on accurate weather information to optimize their operations.

## ****Proposed Solution Approach****

To achieve this, the project uses a multi-step approach that begins with data preprocessing, including the removal of inconsistencies and missing data, followed by feature engineering and scaling. The predictive model is built using a Long Short-Term Memory (LSTM) network due to its strength in time-series forecasting. The model is trained on historical weather data and evaluated for accuracy.

## ****Research Questions and Objectives****

The primary research question is: **Can machine learning techniques, specifically LSTM networks, be used to reliably predict weather conditions based on historical data?**

## The objectives include:

* Developing a machine learning-based forecasting model.
* Ensuring model scalability and robustness.
* Evaluating performance against traditional forecasting methods.

## ****High-Level Results and Contributions****

The model successfully forecasts temperatures, humidity, rainfall, wind speed, and other critical weather conditions. Contributions include:

* Developing a streamlined, automated weather prediction process.
* Handling missing data effectively.
* Presenting a scalable and reusable solution.

# ****Project Management****

A Scrum-based iterative development approach was adopted. The project included a product backlog comprising tasks such as data cleaning, feature engineering, model training, evaluation, and result visualization.

Artifacts:

* **Project Backlog**: Tracked all development tasks and updates.
* **Burn-down Chart**: Illustrated progress over time across sprints, showing completed tasks versus pending ones for each iteration.

# ****Organization****

## ****System Architecture****

The LSTM model's architecture consists of:

* **Input Layer**: Receives time-series weather data.
* **LSTM Layers**: Capture temporal dependencies in data.
* **Dense Layer(s)**: Process data between LSTM and final predictions.
* **Output Layer**: Predicts key weather parameters.

The project was implemented as a standalone application. The architecture is modular and can be extended to support distributed systems if required.

# ****Methods****

## ****Data Collection and Preprocessing****

* **Dataset**: Historical weather data including temperature, humidity, rainfall, wind speed, etc.
* **Cleaning**:
  + Removed invalid or illogical entries.
  + Filled missing values using mean of region and month.
  + Removed outliers.
* **Transformation**:
  + Encoded categorical variables.
  + Converted date fields to datetime format.

## ****Feature Engineering****

* Extracted **Month**, **Year** from date.
* Added **Lag Features** for time-series dependencies.
* Calculated **Rolling Statistics** (mean, std. deviation).

## ****Model Building****

* **Model**: LSTM neural network.
* **Training**:
  + **Loss Function**: Mean Squared Error (MSE)
  + **Optimizer**: Adam
  + **Tuning**: Batch size, learning rate, etc.

## ****Evaluation Metrics****

* Root Mean Squared Error (RMSE)

## ****Tools Used****

* **Python**, with libraries:
  + Pandas, NumPy
  + Scikit-learn
  + TensorFlow/Keras
  + Matplotlib, Seaborn

### ****Learning Experience****

* **Prior Knowledge**: Experience with Python, data analysis, and ML models.
* **New Skills**: Gained experience in LSTM networks, time-series forecasting, and hyperparameter tuning.

### ****Alternatives Considered****

* Considered Random Forests and Gradient Boosting Machines.
* Chose LSTM due to its superior handling of sequential data.

# ****Reflection****

### ****Evaluation of Objectives****

* The model met the goal of accurate forecasting.
* The system demonstrates the practical application of ML for real-world weather prediction.

### ****Model Weaknesses****

* Dependent on data quality.
* Potential overfitting.
* Limited features (e.g., missing geographical or satellite data).

### ****False Starts and Adjustments****

* Adjusted model depth to improve accuracy.
* Improved handling of missing categorical data.

### ****Efficiency and Scalability****

* LSTM is resource intensive.
* GPU usage improved training efficiency but increased cost.
* Model is scalable with proper optimization.

### ****Future Work****

* Incorporate more features (e.g., satellite data).
* Try ensemble models.
* Enable real-time forecasting via streaming data.

# ****Conclusion****

This project demonstrates the viability of using LSTM-based machine learning for weather prediction. While the model is effective, improvements are needed in data handling and computational efficiency. Future iterations may focus on expanding data sources and optimizing for real-time prediction to serve weather-sensitive industries more effectively.

# ****Appendices****

**Code Repository**: <https://github.com/bharathreddyjakkidi/Capstone-G02520049>

**Sample Output Visualizations**

