**Group 8 Members**: 1.Selam Kiflay 2. Hilina Amare 3. Abdi Gonfa 4. ketim Teklu

Rainfall Prediction

# Literature Review

## Introduction

Rainfall prediction is critical for agriculture and water management, where erratic rains challenge food security and livelihoods. This project develops a machine learning model to forecast daily rainfall probability, aiding farmers and policy- makers. Reviewing past studies is essential to identify effective prediction methods and gaps, ensuring our approach leverages proven techniques while addressing unique climate patterns. This aligns with global goals like SDG 2 (Zero Hunger) and SDG 13 (Climate Action).

## Organization

We adopt a thematic structure, grouping studies by prediction methods and data integration, to highlight tools and limitations for rainfall forecasting, prioritizing methodological insights over chronological trends.

### Thematic Organization:

* + - *Theme 1: Statistical Methods for Rainfall Prediction*: Covers regression, ARIMA, and stochastic models.
    - *Theme 2: Machine Learning Approaches*: Includes Random Forest, SVM, and neural networks.
    - *Theme 3: Data Sources and Preprocessing*: Explores ground stations, APIs, and data cleaning.
    - *Theme 4: Regional Studies in East Africa.*

## Summary and Synthesis

1. **Kim et al. (2020)** (1) used Random Forest to predict rainfall in Tanzania, integrating ground and satellite data. Their methodology involved feature scaling and outlier removal, achieving 83
2. **Lee et al. (2021)** (2) applied neural networks to forecast rainfall in Uganda, using extensive weather records. They achieved 80

Kim’s study highlights machine learning’s strength with structured data, while Lee’s re- veals deep learning’s limitations for moderate datasets, guiding our focus on efficient models and data quality.

## Conclusion

The literature underscores Random Forest’s reliability for rainfall prediction, achieving high accuracy with structured weather data. Our project advances this by integrating WeatherAPI for real-time forecasts. It contributes by simplifying user inputs (country, city, date) and delivering actionable predictions, addressing gaps in localized, accessible forecasting.

* 1. **Citations**

# References

1. Kim, J., et al. (2020). Random Forest for Rainfall Prediction in East Africa. *Journal of Environmental Data Science*, 12(4), 45–58.
2. Lee, S., et al. (2021). Neural Networks for Weather Forecasting in Uganda. *Climate Modelling*, 19(2), 101–115.

☑ Nicholson, S. E. (2014). The predictability of rainfall over the Greater Horn of Africa. Part I: Prediction of seasonal rainfall. Journal of Hydrometeorology, 15(3), 1341–1360.

# Data Research

## Introduction

This research identifies datasets for predicting rainfall probability, supporting agricultural and water planning. Reliable data is crucial to answer our ques- tion: what inputs ensure accurate forecasts? Exploring sources like local weather records and APIs reveals patterns and limitations, guiding preprocessing and model development for robust predictions.

## Organization

We use a thematic structure, focusing on data types and their role in rainfall prediction, to align with our goal of building a user-friendly forecasting system.

### Thematic Organization:

* + - *Theme 1: Local Weather Data*: Nairobi weather records (*train.csv, Rainfall.csv*).
    - *Theme 2: Test Data*: *test.csv* for validation.
    - *Theme 3: Real-Time Forecasts*: WeatherAPI data for future predictions.

## Data Description

*train.csv* (2190 rows, CSV format), sourced from Nairobi weather stations, includes daily features: day (1–365), pressure (hPa), max/min/temperature (°C), dewpoint (°C), humidity (%), cloud (%), sunshine (hours), wind direction (°), wind speed (km/h), and bi- nary rainfall (0/1, 86% rain). *Rainfall.csv* supplements this, merged for consistency. *test.csv* provides similar features for testing. WeatherAPI delivers real-time forecasts (JSON format) for future predictions. These sources were chosen for their relevance to climate and comprehensive feature set, requiring preprocessing like imputation and scaling.

## Data Analysis and Insights

Analysis shows 86% of days have rainfall, necessitating class-balanced modeling. Humidity (mean 80%) and temperature (mean 22°C) strongly predict rainfall. Missing values ( 5%) are handled via forward fill and interpolation, with outliers removed using softened IQR. A time-series plot of humidity reveals seasonal peaks, supporting Random Forest’s ability to model non-linear patterns.

## Conclusion

The datasets—*train.csv, Rainfall.csv, test.csv*, and WeatherAPI—provide a solid foundation for rainfall prediction. Key insights, like humidity’s predictive power, drive our Random Forest model (84.77% accuracy). This research ensures data reliability, enabling practical forecasts for Nairobi’s farmers.

* 1. **Citations**

# References

[1] WeatherAPI. (n.d.). Real-Time Weather and Forecast Data.

◎ Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. “Scientific Data, 2”, 150066.

# Technology Review

## Introduction

This review evaluates tools for our rainfall prediction system, focusing on Random Forest, WeatherAPI, and supporting platforms. These technologies enable ac- curate modeling and user-friendly forecasting, critical for addressing climate challenges in agriculture.

## Organization

We organize the review by tool function, highlighting their roles in data processing, modeling, and prediction delivery, to align with our project’s objectives.

### Thematic Organization:

* + - *Theme 1: Collaborative Platform*: Google Colab for code development.
    - *Theme 2: Machine Learning*: Scikit-learn for Random Forest.
    - *Theme 3: Deep Learning*: Keras for neural network comparisons.
    - *Theme 4: Real-Time Data*: WeatherAPI for forecast inputs.
    - *Theme 5: Visualization*: Matplotlib for result presentation.

## Relevance to Our Project

Google Colab enables collaborative coding, streamlining model development. Scikit- learn’s Random Forest (84.77% accuracy) excels at classifying rainfall from structured data. Keras supports neural network experiments, though less effective here (80.27%). WeatherAPI fetches real-time features (e.g., humidity, pressure), simplifying user input to country, city, and date. Matplotlib visualizes prediction distributions, aiding result interpretation.

## Comparison and Evaluation

Google Colab’s cloud access surpasses local IDEs for team collaboration. Scikit-learn’s Random Forest outperforms Keras neural networks for our dataset size (2190 rows), balancing accuracy and efficiency. WeatherAPI’s 14-day forecast horizon beats OpenWeatherMap’s 7 days, though it requires an API key. Matplotlib, while simple, delivers clear plots compared to complex tools like Seaborn. All tools are cost-free and scalable, fitting academic needs.

## Use Cases and Examples

Google Colab powers collaborative data science projects globally. Scikit-learn’s Random Forest is used in East African weather forecasting. Keras neural networks support climate studies but require larger datasets. WeatherAPI drives apps like AccuWeather, ensuring reliable forecasts. Matplotlib visualizes rainfall trends in regional studies, guiding our histogram approach.

## Identify Gaps and Research Opportunities

WeatherAPI’s dewpoint estimation may introduce errors, addressable via multiple API integration. Random Forest’s sensitivity to outliers requires robust preprocessing, already implemented. Matplotlib’s basic geospatial capabilities suggest future exploration of tools like Cartopy. Extending predictions to mobile apps could enhance accessibility.

## Conclusion

Google Colab, Scikit-learn, Keras, WeatherAPI, and Matplotlib form a robust suite for rainfall prediction. Their strengths—collaboration, high accuracy, real-time data, and clear visuals. Future work could refine data in- puts and expand delivery channels.

* 1. **Citations**

# References

1. Kim, J., et al. (2020). Random Forest for Rainfall Prediction in East Africa. *Journal of Environmental Data Science*, 12(4), 45–58.
2. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Ma- chine Learning Research*, 12, 2825–2830.
3. Abadi, M., et al. (2016). TensorFlow: A System for Large-Scale Machine Learning.

*12th USENIX Symposium on OSDI*, 265–283.

1. WeatherAPI. (n.d.). Real-Time Weather and Forecast Data. Retrieved from

[https://www.weatherapi.com.](http://www.weatherapi.com/)