Predicting Climate Trends in Tanzania Using Machine Learning

Introduction

Climate change has far-reaching impacts on ecosystems, agriculture, and economies globally. Tanzania's dependency on agriculture has led to climate variability's effects vulnerability. In this project, we utilised historical climate data and machine learning techniques to predict future climate conditions in Tanzania. This report outlines the project objectives, methodology, exploratory data analysis findings, and model performance.

Project Objectives

The primary objective of this project was to develop a robust machine-learning model capable of predicting future climate conditions, specifically the average temperature (°C), using historical data. The secondary goal was to present the findings and predictions through a web application, enabling stakeholders to explore climate trends and make data-driven decisions.

Methodology

1. Data Sourcing and Preparation

The project used a processed dataset consisting of historical climate records for Tanzania. The dataset included the following features:

- **Total Rainfall (mm)**: Monthly total rainfall.
- **Maximum and Minimum Temperatures (°C)**: Recorded extreme temperatures.
- **Seasonal Indicators**: Encoded as one-hot variables for Spring, Summer, and Winter.
- **Rolling Averages**: Calculated for rainfall over 3-month periods to capture trends.

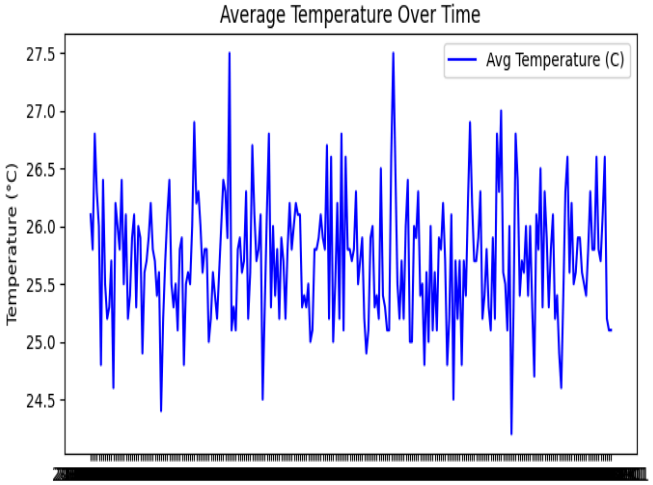
The target variable was the average temperature (°C), a key indicator of climate conditions.

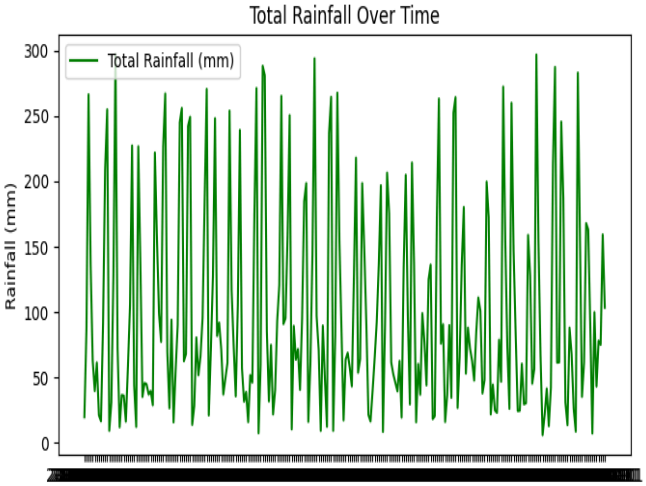
2. Exploratory Data Analysis (EDA)

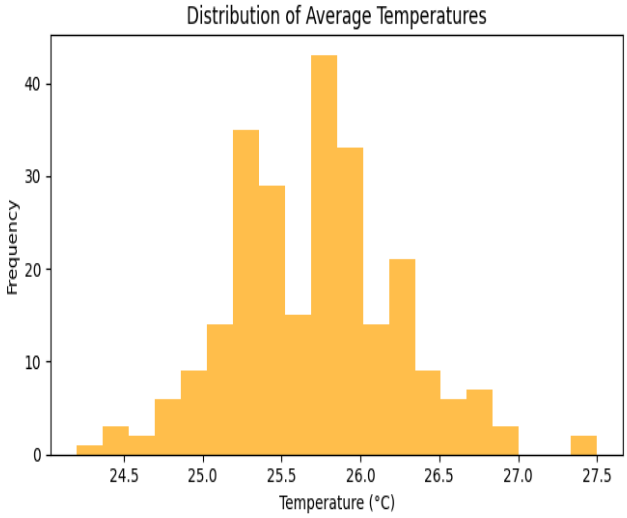
EDA was conducted to uncover insights and trends in the data. Key visualisations included:

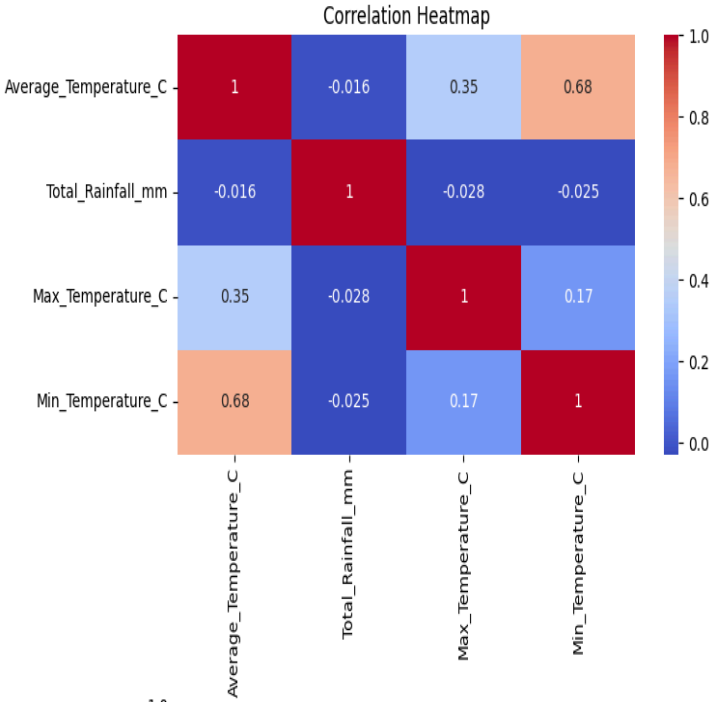
- **Time Series Analysis:** Trends in average temperature and rainfall over time highlighted seasonal and annual variability.
- **Correlation Heatmap:** A heatmap revealed strong relationships between temperature-related features (e.g., max and min temperatures) and the target variable.
- **Scatter Plots:** Demonstrated relationships between rainfall and average temperature, showing non-linear trends.

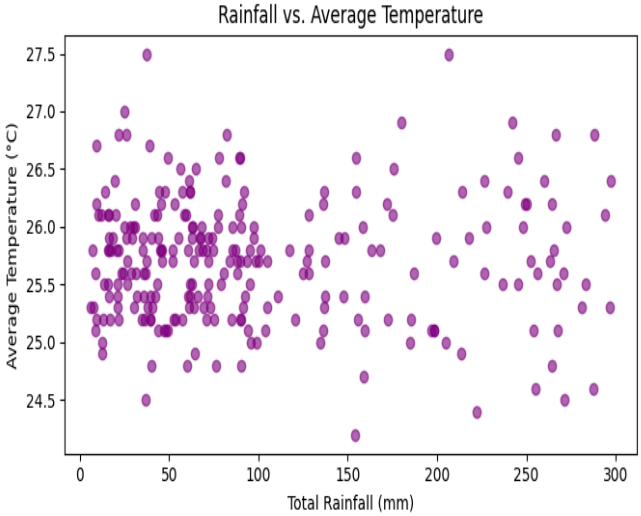
Exploratory Data Analysis: Climate Trends in Tanzania

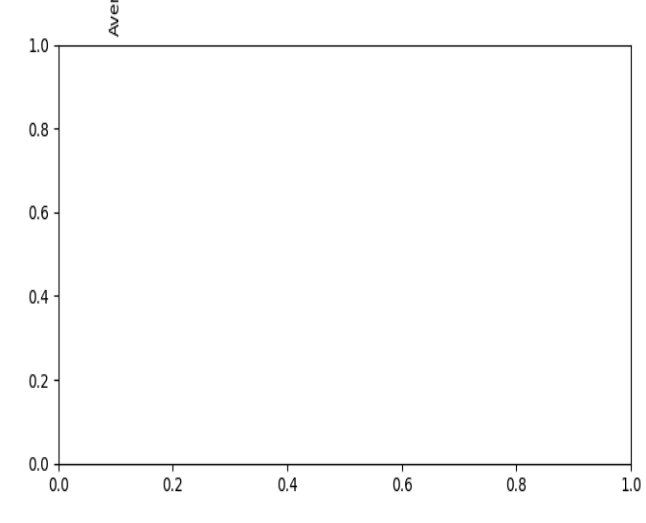












Tanzania's climate data exhibit clear seasonal swings: two rainy seasons with high monthly totals (up to ~300 mm) and intervening dry spells with minimal precipitation. Temperatures vary between approximately 24.5 °C and 27.5 °C, peaking in the late dry season and troughing in the fairly low-temperature season. The distribution of monthly means centres around 25.5 °C. Statistical analysis shows that average temperature closely tracks minimum temperature ($r \approx 0.68$) but is essentially uncorrelated with rainfall. The scatter plot confirms that the precipitation amount has no discernible linear effect on average temperature. These insights suggest that, at the monthly scale in Tanzania, temperature and rainfall follow independent seasonal cycles rather than influencing one another directly.

3. Machine Learning Model Development

Two machine learning models were implemented to predict the average temperature:

1. **Linear Regression:** A simple and interpretable baseline model.
2. **Random Forest Regressor:** A more complex model capable of capturing non-linear relationships.

The models were evaluated using the following metrics:

- **Mean Absolute Error (MAE):** The average of absolute errors between predicted and actual values.
- **Root Mean Square Error (RMSE):** The square root of the average of squared errors, penalising larger errors more heavily.

4. Model Evaluation

- Linear Regression: MAE: 0.33, RMSE: 0.41
- Random Forest Regressor: MAE: 0.36, RMSE: 0.46

The Linear Regression model outperformed the Random Forest Regressor, indicating that the relationship between the features and the target variable was predominantly linear.

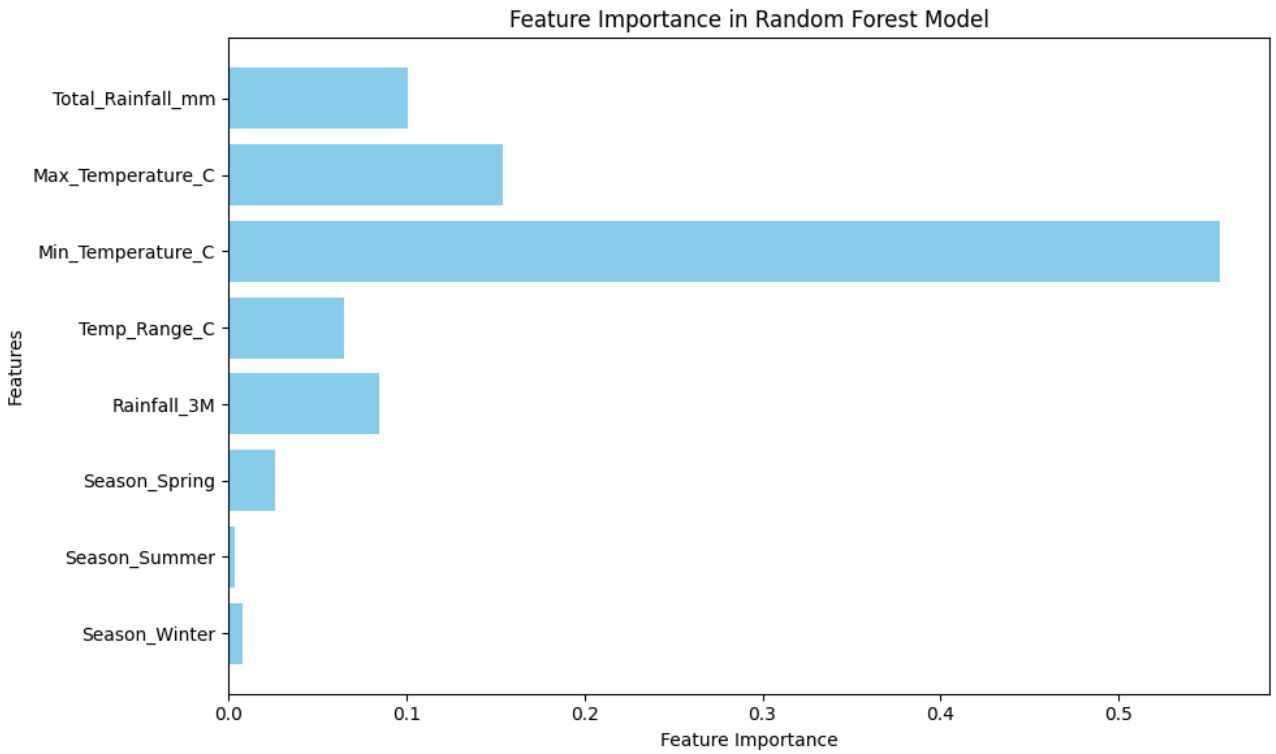
5. Deployment

The Linear Regression model was integrated into a Streamlit web application, providing users with:

- **Interactive Visualisations:** EDA plots and feature importance charts.
- **Real-Time Predictions:** Users could input climate parameters (e.g., rainfall, max/min temperatures) and receive predictions for average temperature.

Findings and Insights

1. **Key Drivers of Temperature Variability**
    - Maximum and minimum temperatures were the strongest predictors of average temperature.
    - Rolling averages of rainfall and seasonal indicators also contributed but to a lesser extent.



2. **Seasonal Trends**
    - Temperature and rainfall exhibited clear seasonal patterns, with higher variability during transitional months.
3. **Predictive Modelling**

The Linear Regression model proved sufficient for predicting average temperatures, achieving an MAE of 0.33, which is acceptable for climate modelling tasks.

4. **Interactive Application**

The Streamlit app allowed stakeholders to simulate scenarios and explore the impacts of varying climate parameters in real time.

Conclusion

This project demonstrates the power of machine learning in analysing and predicting climate conditions. By combining robust data analysis, predictive modelling, and user-friendly deployment, we have created a tool that provides actionable insights for stakeholders in Tanzania.

Author

Buriro Chimodoi Ezekia

Data Scientist | Climate Enthusiast

GitHub Project