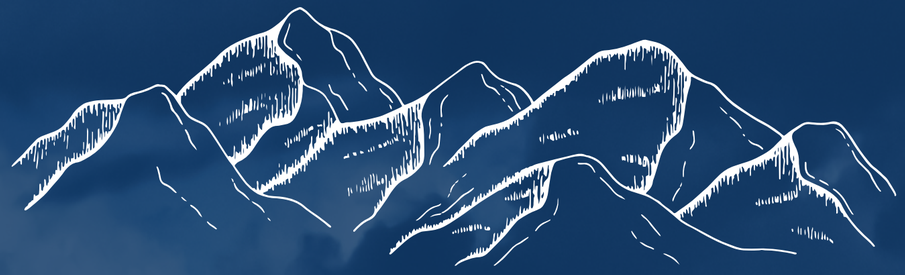

RAINFALL PREDICTION

GRACE GITAU

OVERVIEW

This project revolvs around leveraging historical weather data to forecast conditions in Nairobi. Key aspects include data preprocessing, encompassing cleaning and feature engineering, to enhance the dataset's quality. Various machine learning algorithms were applied to build a robust predictive model, focusing on rain prediction. Evaluation metrics and validation techniques ensured model effectiveness, helping those sectors which were most dependent on weather, such as agriculture.

BUSINESS PROBLEM

The challenge at hand is to improve the accuracy of rainfall predictions for Nairobi, utilizing a decade's worth of historical weather records. Current prediction methods often fall short of providing precise forecasts tailored to Nairobi's unique climatic conditions, emphasizing the need for a robust predictive model.

DATA UNDERSTANDING

The dataset was sourced from Kaggle and includes various weather attributes such as Temperature, Rainfall, Humidity, Wind Speed, and more. The data covers the period from 1970 to 1999.

OBJECTIVES

1. Pre-process and analyze the extensive weather data.

2. Develop, train and validate machine learning models for accurate rainfall prediction.

3. Assess model performance and optimize parameters to improve accuracy & reliability.

MODELING

The project utilized three primary models:

1. **Logistic Regression**
2. **Random Forest**
3. **XGBoost**

Feature engineering and preprocessing were applied to the dataset, followed by training the models. Each model's performance was evaluated using accuracy, precision, recall, and F1-score.

EVALUATION AND CONCLUSION

Among the models tested, XGBoost outperformed others in terms of accuracy, precision, recall, and F1-score, making it the most effective model for predicting rainfall in this context.

Random Forest also showed strong performance but exhibited some overfitting, as indicated by the high training accuracy and slightly lower testing accuracy. Logistic Regression, while less complex, provided a good balance between computation time and accuracy.

RECOMMENDATIONS

Model Deployment:

01 The XGBoost model, due to its superior performance, should be considered for deployment in real-time applications. Integration with local weather stations and government platforms could enhance weather prediction accuracy and provide timely alerts to farmers and other stakeholders.

Data Expansion:

02 To further improve the model's accuracy, it is recommended to incorporate more recent weather data, possibly expanding beyond the current dataset. Additionally, integrating satellite data and other environmental variables could capture more complex patterns in rainfall prediction.

Regular Model Updates:

03 The model should be periodically retrained and validated with new data to ensure its accuracy remains high. Regular updates will help the model adapt to changing climatic conditions and provide more reliable forecasts over time.

Further Research:

04 It is advisable to explore other advanced algorithms and techniques, such as deep learning models, which might offer improvements in handling large datasets with complex relationships, potentially leading to even more accurate predictions.

Thank You

grace.gitau2@student.moringaschool.com