**CROP PRICE PREDICTION**

**INTRODUCTION**

Datawise Solutions specializes in providing cutting-edge data analytics and consulting services to help enterprises of all sizes maximize the value of their data. With a team of expert data scientists, analysts, and consultants, we are dedicated to supporting organizations in gaining valuable insights, making informed decisions, and driving growth.

Our goal is straightforward: convert raw data into useful intelligence. Whether you want to streamline operations, better marketing techniques, enhance customer experiences, or drive innovation, we offer the expertise and technology to help you turn data challenges into opportunities.

From data collection and cleansing to advanced analytics and predictive modeling, we offer a comprehensive suite of services adapted to your specific needs. Our collaborative approach ensures that we grasp the exact business objectives and create solutions that produce demonstrable outcomes. Datawise Solutions believes that the power of data is infinite. Let us be your reliable partner on the path to data-driven success.

**Business Understanding.**

According to a Business Daily article, Kenyan farmers earn less even though consumers pay more for food products as there is a huge disparity between farmers' price and consumer prices. Kenyan Farmers commodities prices are influenced by the inefficient market structure and the intermediaries between producers and consumers. It is noted that the market dynamics are dictated by the middlemen contributing to divergent wholesale and retail prices in the Kenyan Markets. However, there are external factors that have affected farmers' commodities prices such as subsidy schemes, currency fluctuation but farmers continue to earn less than the market price set in Kenyan stores and marketplaces.

According to the KARI report, the supply side factors, there is a decline in agricultural productivity, due to poor market access and market exploitation. For small-scale farmers Agriculture is a key contributor to the Kenyan GDP and provides livelihoods to a large part of the Kenyan Population. CBK Survey 2023 shows that the food commodities prices are down. In addition, the survey shows that foreign currency performance could be a factor in downward adjustment of the food prices.

For optimal business focus, we consider prioritizing the following crops: Maize, Beans, Onions, Wheat, Rice. These crops offer potential for profitability and market demand, making them strategic choices for your agricultural ventures and market exploitation. For small-scale farmers. Agriculture contributes significantly to Kenya's GDP and employs a big proportion of the Kenyan population.

The CBK Survey 2023 shows that food commodity prices are falling. Furthermore, the poll reveals that foreign currency performance may play a role in food price reductions. For optimal business emphasis, we recommend emphasizing the following crops: maize, beans, onions, wheat, and rice. These crops have the potential for profitability and market demand, making them excellent selections for your agricultural ventures.

**Research Question.**

1. What are the key factors influencing price disparities between farmers' earnings and consumer prices in the Kenyan agricultural sector?
2. How effective are predictive pricing updates in enabling farmers to make informed decisions regarding resource allocation and market positioning?
3. What role do government agencies, financial institutions, and other stakeholders play in promoting resilience and sustainability within the Kenyan agricultural sector?
4. How can collaborative partnerships between stakeholders, including farmers, government agencies, and financial institutions, drive positive change within the Kenyan agricultural sector?

**Problem Statement.**

In Kenya, farmers often face significant disparities between the prices they receive for their crops and the prices consumers pay at marketplaces. This gap is primarily driven by inefficient market structures and the presence of intermediaries who exploit the lack of transparency in pricing. As a result, farmers struggle to earn fair prices for their produce, despite being a crucial component of the country's GDP and a source of livelihood for many. To address this challenge, there is a pressing need for a comprehensive price forecasting system that leverages historical and real-time data, supported by machine learning models. Such a system would empower farmers with predictive pricing updates, enhance market transparency, and enable informed decision-making for all stakeholders along the agricultural value chain.

**Objectives.**

* To enable farmer's and other stakeholders to anticipate the market trends of the crop produce, resource allocation, and market position
* To enhance market transparency and price forecasting to farmers and industry policymakers.
* To predict the price food products based on supply and demand of food products and market location
* Market segmentation for the farmers based on the prices, supply volumes, and location. - Best Market Location Place and Product to sell.

**Data Understanding.**

The dataset contains 14 columns and 1271 rows. The columns include the following:

**Date:** The date of the exchange rate data.

**Price**: The closing price of USD to KES exchange rate on that date.

**Open**: The opening price of USD to KES exchange rate on that date.

**Low**: The lowest price of USD to KES exchange rate on that date.

**Volume**: Volume of USD to KES exchange transactions (seems to be mostly NaN values).

**Change %**: Percentage change in the USD to KES exchange rate from the previous day.

**Product**: Type of food product.

**Market Location**: Location of the market where the product is sold.

**Wholesale Price**: Wholesale price of the product.

**Supply Volume**: Volume of supply for the product.

**County**: County where the market is located.

**Date**: Date of the record.

**USD Rate**: USD exchange rate.

**Additional engineered features:** Ratio of supply volume to wholesale price, one-hot encoded product and market location, date-time features, lag features, rolling window statistics, and differenced features.

After merging the two datasets our data frame contains 61,165 entries. Here's a summary of the data frame:

Columns: Product, Market Location, Wholesale Price, Supply Volume, County, Date, USD rate

This dataset is used to analyze the trends in USD to KES exchange rates over time, which can be helpful in understanding currency fluctuations and their impact on the prices of agricultural commodities considering other factors too.

**Data preparation**

**Loading the data.**

After that, the datasets were added to the Jupyter Notebook, where it was previewed for a better understanding of their columns and the relationships that exist between them.

**Cleaning data.**

Data cleaning was performed from the data frame and the following columns were dropped Open, High, Low, Vol. and Change %. Additionally, the column price was renamed USD rate. The dataframe was scrutinized to check for missing values and all missing values were handled. After handling missing values, the dataset no longer contains any missing values.

**Feature Engineering.**

Feature engineering was done and price column was renamed usd\_rate. The Ratio of Supply Volume to Wholesale Price was done. The year, month, and day from the Date column was extracted and the original date column was dropped. One hot encoding was done on the categorical columns and finally checked for missing values in the encoded dataset and it was discovered that there were no missing values.

**EDA**

Exploratory Data Analysis (EDA) served very important roles like understanding the data. The visualizations helped to identify several features like outliers. Visualizations like scatterplots, bar graphs, boxplots and line chart gave insights. The line chart depicted the wholesale prices of Beans, Dry Maize, Onions and Rice over time. On the other hand, scatterplots visualized relationships between agricultural product prices, supply volumes, and USD exchange rates. Bar plots revealed the comparison of average wholesale price by market location and the comparison of supply volume by market location.

**Modelling**

Our project commenced with an in-depth exploration of the dataset using **Linear Regression**. The model produced the following metrics: RMSE: 31.12, MSE: 968.35. These metrics provided a baseline for assessing predictive performance and guided subsequent modeling efforts.

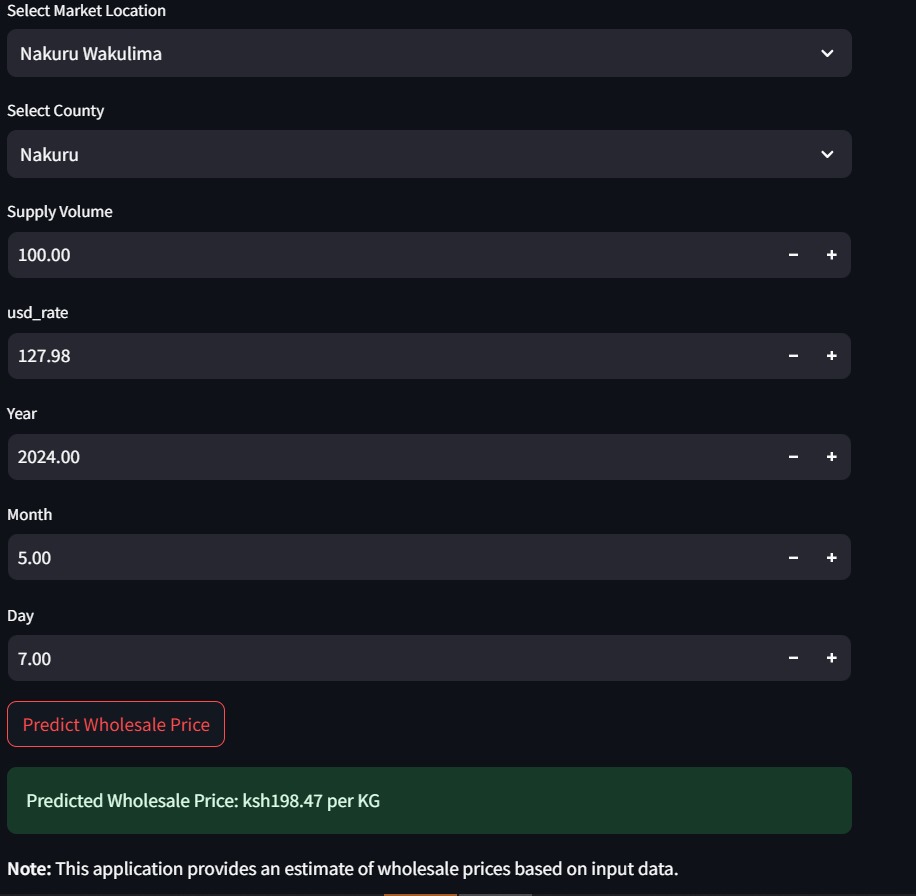
Following the initial Linear Regression analysis, our project progressed to employing **Random Forest modeling**. The Random Forest model yielded the following performance metrics: RMSE: 9.75, MSE: 95.11. These results indicated a significant improvement in predictive accuracy compared to the Linear Regression model.

We advanced to utilizing **XGBoost model** which yielded this performance metrics: RMSE: 6.79, MSE: 46.17. These results demonstrated further enhancement in predictive accuracy compared to both Linear Regression and Random Forest models.

In the final phase of our analysis, we implemented a **Decision Tree model** and it achieved the following performance metrics: RMSE: 12.23, MSE: 149.49. Comparing the performance of all models, XGBoost emerged as the best-performing model with the lowest RMSE and MSE values. Additionally, XGBoost attained a score of 0.94, indicating its strong predictive capability. Therefore, XGBoost outperformed Linear Regression, Random Forest, and Decision Tree models, showcasing its effectiveness in this analysis.

**User interaction**

For user interaction we designed a user interface that incorporates all metadata that are responsible in predicting wholesale prices in various market locations. The user accesses the deployment interface and keys in product, location, county, the date and populates the wholesale price of the products



**Conclusions**

In conclusion, this project has successfully achieved its objectives of developing a comprehensive price forecasting system for the Kenyan agricultural market. Through the implementation of machine learning models and strategic analysis, we have empowered farmers with predictive pricing updates and enhanced market transparency. Our focus on strategic crops has provided valuable insights into market dynamics and opportunities for growth. Moving forward, we are poised to continue driving positive change within the agricultural sector, promoting resilience and sustainability for farmers and stakeholders alike.

**Recommendations**

Based on our analysis and findings, we recommend the following actions:

**1.Implementation of Predictive Pricing System:** Deploy the developed machine learning models to provide farmers with predictive pricing updates, enabling them to make informed decisions regarding resource allocation and market positioning.

**2.Enhanced Market Transparency:** Collaborate with government agencies and agricultural traders to promote market transparency and fair pricing practices, thereby reducing the influence of intermediaries and ensuring farmers receive fair prices for their produce.

**3.Continuous Monitoring and Evaluation:** Regularly monitor and evaluate the performance of the price forecasting system to ensure accuracy and relevance. Incorporate feedback from stakeholders to improve the system's effectiveness over time.

**Future improvement ideas**

To further enhance the effectiveness and impact of the price forecasting system, we suggest exploring the following areas for future improvement

1. Integration of Real-Time Data: Incorporate real-time data sources, such as weather forecasts and market trends, to enhance the accuracy and timeliness of price predictions.
2. Expansion of Market Segmentation Analysis: Expand market segmentation analysis to provide farmers with insights into the best market locations and products to sell, based on pricing, supply volumes, and location dynamics.
3. Capacity Building and Training: Provide training and support to farmers in utilizing the price forecasting system effectively, ensuring widespread adoption and empowerment across the agricultural community.

**Deployment**

In today's dynamic agricultural landscape, access to timely and accurate market information is crucial for farmers to make informed decisions and optimize their livelihoods. Recognizing this need, the Kenya Farmers Association (KFA) is taking a significant step forward by integrating wholesale price prediction tools into its website. This strategic initiative aims to empower farmers and stakeholders across Kenya with actionable insights to navigate market uncertainties and enhance their economic prospects.

The integration of wholesale price prediction tools into the KFA website represents a transformative leap towards empowering farmers with real-time market intelligence. Leveraging factors such as the USD rate, location, county, product type, and supply volume, these tools provide invaluable insights into wholesale price dynamics specific to the Kenyan agricultural market. Farmers can now anticipate price trends, plan their crop production, and negotiate better prices with buyers, ultimately improving their profitability and livelihoods.

One of the key advantages of these prediction tools is their ability to offer real-time forecasts, enabling farmers to make timely and informed decisions regarding crop sales and marketing strategies. By accessing up-to-date price predictions through the KFA website, farmers can optimize their selling timing, minimize market risks, and maximize their returns from agricultural produce. This not only enhances the financial well-being of individual farmers but also strengthens the overall resilience and competitiveness of the agricultural sector.